

# **Series enlazadas de algunos agregados económicos nacionales y regionales, 1955-2007**

**Versión 2.1**

**Angel de la Fuente\***

**Instituto de Análisis Económico (CSIC)**

**Septiembre de 2009**

## **Resumen**

En este trabajo se construyen series “homogéneas” de VAB a precios corrientes y constantes, empleo y población para España y sus regiones durante el periodo 1955-2007. Estas series se obtienen enlazando la Contabilidad Regional del INE con las series elaboradas por Julio Alcalde y colaboradores para la Fundación BBVA. El “punto de corte” en el que se abandona esta última fuente en favor de la anterior como referencia para la construcción de la serie enlazada se determina utilizando un procedimiento que permite estimar cuál de las dos series disponibles generará un estimador con menor error cuadrático medio cuando se utiliza como variable dependiente en una regresión sobre una variable independiente arbitraria. La discrepancia entre ambas series que aflora en el momento del enlace se reparte entre los niveles iniciales y las tasas de crecimiento de la serie más antigua utilizando, en la medida de lo posible, estimaciones externas del valor de las variables de interés a comienzos del período muestral.

## **Abstract**

I construct “homogeneous” series of GVA at current and constant prices, employment and population for the Spain and its regions covering the period 1955-2007. The series are obtained by linking the Regional Accounts of the National Statistical Institute with the series constructed by Julio Alcaide and his team for the BBVA Foundation. The “switching point” at which this last source stops being used as a reference to construct the linked series is determined using a procedure that allows me to estimate which of the two competing series would produce an estimator with the lowest MSE when it is used as dependent variable in a regression on an arbitrary independent variable. To the extent that it is possible, the difference between the two series found at the point of linkage is distributed between the initial levels of the older series and its subsequent growth using external estimates of the relevant variables at the beginning of the sample period.

---

\* Este trabajo forma parte de un proyecto de investigación cofinanciado por el Fondo Europeo de Desarrollo Regional y la Fundación Caixa Galicia. Agradezco también la financiación del Ministerio de Ciencia e Innovación a través del proyecto ECO2008-04837/ECON y los comentarios, sugerencias y ayuda con los datos de Antonio Díaz, Andrés de Bustos, Javier Escribá y Alfredo Cristóbal.

## 1. Introducción

Una de las dificultades con las que nos enfrentamos los analistas de la economía regional española es la inexistencia de series homogéneas comúnmente aceptadas de los principales agregados económicos regionales para períodos largos de tiempo. Las únicas series existentes que se remontan a 1955 son las elaboradas por Julio Alcaide y diversos colaboradores, primero para el Servicio de Estudios del Banco de Bilbao y posteriormente para la Fundación BBVA y la Fundación FIES. Por otro lado, disponemos también de la Contabilidad Regional de España (CRE) que elabora el Instituto Nacional de Estadística (INE) desde 1980. La CRE ha sufrido importantes cambios metodológicos a lo largo del tiempo, sin que esté disponible hasta el momento una serie histórica homogénea oficial para el conjunto del período. Ambas fuentes, además, presentan diferencias muy apreciables durante su período de solapamiento.

En el presente trabajo se construyen series históricas "homogéneas" de algunos agregados regionales (VAB a precios corrientes y constantes, empleo y población). Las nuevas series se construyen enlazando la Contabilidad Regional en base 2000 (CRE00) con series anteriores de la propia CRE y con los datos de la Fundación BBVA.<sup>1,2</sup>

El trabajo está organizado como sigue. En la sección 2 se describe la información disponible sobre la evolución de los agregados de interés. En la sección 3 se discuten los principales problemas metodológicos que se plantean en la construcción de series enlazadas a partir de esta información, así como la forma en que éstos se han abordado en el estudio. El primero de estos problemas tiene que ver con la elección de la serie de referencia en aquellos períodos en los que hay varias disponibles o, en términos prácticos, con la fijación de la "fecha de corte" en la que se abandonan las series del BBVA para adoptar como referencia los datos de CRE elaborados por el INE. En el primer apartado de la sección 3 se desarrolla una metodología que permite identificar (y adoptar como referencia) la serie con mayor contenido informativo. El procedimiento de selección se base en el siguiente experimento conceptual: supongamos que queremos analizar los determinantes de cada una de las variables con las que estamos trabajando (renta y empleo) regresándola sobre alguna otra variable,  $x$ . ¿Cuál de las series

---

<sup>1</sup> La idea de enlazar las series de la CRE con las del BBVA no es novedosa. Doménech et al (1999) y de Bustos et al (2008) realizan sendos ejercicios en esta línea. En ambos trabajos se toman como dadas las series de la CRE desde su inicio en 1980 y se extienden hacia atrás tomando como referencia las series del BBVA. Doménech et al (1999) ofrecen dos enlaces alternativos entre ambas fuentes: el primero de ellos se construye por retropolación (esto es, las tasas de crecimiento de las series del BBVA se utilizan para prolongar hacia atrás las series de la CRE), mientras que en el segundo se respetan los valores iniciales de la serie más antigua, repartiéndose a lo largo del período las diferencias entre ambas fuentes que afloran en 1980. De Bustos et al (2008) optan por el primero de los métodos citados. En ambos casos, las series así obtenidas se ajustan de forma que su agregación coincida con una serie nacional de referencia basada en la CNE. En ambos trabajos, el enlace se realiza con datos desagregados por sectores (cuatro grandes ramas en Doménech et al y 17 en de Bustos et al). Las principales diferencias entre este trabajo y los dos citados tienen que ver con el nivel de agregación, con el método de enlace y con la determinación del "punto de corte" entre las series del BBVA y de la CRE. Los dos últimos aspectos se discutirán en detalle más adelante.

<sup>2</sup> Los cambios más importantes que se introducen en este trabajo en relación con su versión anterior (de la Fuente, 2008) son dos. En primer lugar, las series terminales son distintas (entre 1995 y 1999). Mientras que en de la Fuente (2008) se tomaba como dado el enlace construido por el INE entre las series de la CRE95 y la CRE00, en el presente trabajo se parte del enlace alternativo descrito en de la Fuente (2009). En segundo lugar, el procedimiento de enlace también ha cambiado: en la mayor parte de los casos he abandonado la retropolación en favor del llamado *método mixto*, desarrollado en el presente trabajo (véase la sección 3.2).

disponibles generará un estimador con mejores propiedades y, en particular, con un menor error cuadrático medio?

El segundo problema es la selección del método de enlace. En la segunda parte de la sección 3 se describen brevemente los métodos de enlace más habituales, interpretándolos como formas alternativas de estimar y corregir el error de medición contenido en la serie más antigua, para el que sólo disponemos de la observación correspondiente al momento del enlace. Tras constatar que los procedimientos de enlace habituales imponen hipótesis muy extremas sobre el comportamiento este error, se desarrolla un procedimiento *mixto* que permite una mayor flexibilidad a la hora de repartir las correcciones necesarias entre los niveles iniciales y las tasas de crecimiento de la serie más antigua. En la mayor parte de los casos, éste es el método que utilizaré en el presente trabajo, apoyándome en la medida de lo posible en información externa para fijar el valor de un parámetro que viene a medir la fracción de la discrepancia aflorada en el momento del enlace que ya existía en el año inicial de la serie más antigua.

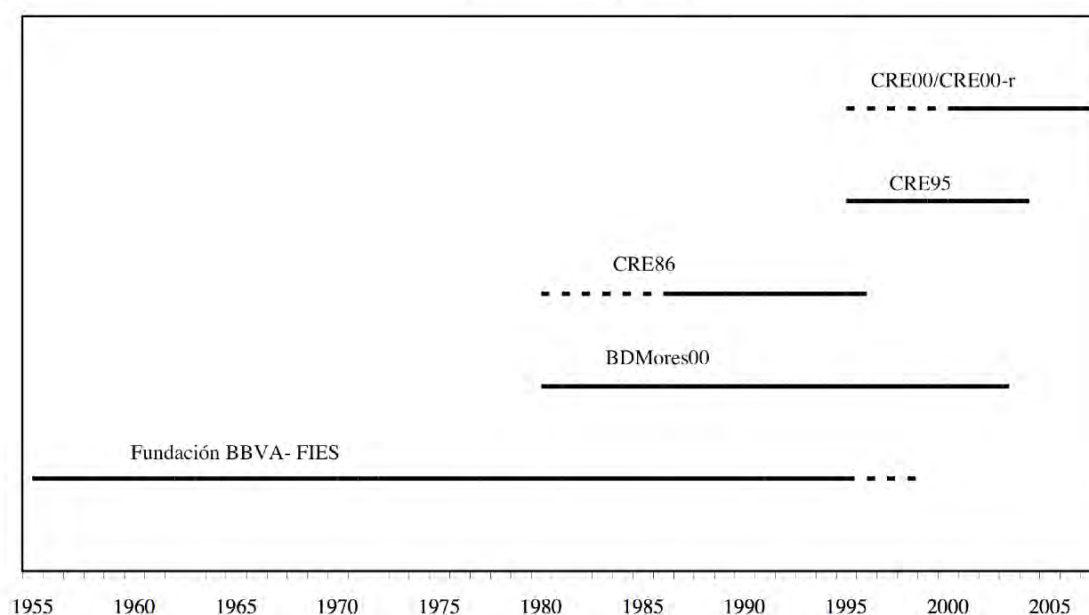
En la sección 4 se construyen series enlazadas de población (de hecho y de derecho) a nivel nacional y regional y en la sección 5 se aborda el enlace del empleo y del VAB a precios corrientes y constantes para España en su conjunto, partiendo fundamentalmente de las sucesivas series de Contabilidad Nacional (CNE). La construcción de las series regionales de estos agregados se describe en la sección 6. El ejercicio se realiza en primera instancia en términos relativos, recuperándose después los niveles de las series a partir de los correspondientes valores nacionales. Esto es, para cada una de las fuentes disponibles, se construyen índices de renta y empleo por habitante, normalizados por el correspondiente promedio nacional. Tras seleccionar la fuente más fiable en cada período por el procedimiento esbozado arriba, se construyen series de referencia de renta y empleo relativos para la parte inicial del período muestral, que se enlazan a continuación con la serie reciente considerada más fiable, normalizada de la misma forma. Una vez enlazados los índices relativos, los niveles per cápita de las series se recuperan multiplicando éstos por los niveles medios nacionales medidos en euros o en puestos de trabajo por habitante contruidos en la sección 5. Finalmente, los valores totales de las series se recuperan multiplicando sus valores per cápita por la población de cada región.

La sección 7 concluye. En el Anexo 1 se describe la construcción de algunas series auxiliares así como los detalles técnicos del procedimiento utilizado para fijar la fecha de corte entre los datos del BBVA y los de la CRE. En el Anexo 2 se incluyen una serie de gráficos en los que las nuevas series enlazadas se comparan con otras ya existentes. Finalmente, el trabajo va acompañado de dos hojas de cálculo. La primera contiene las nuevas series enlazadas y la segunda los datos originales que se han utilizado para construirlas.

## **2. Información disponible**

El Gráfico 1 resume la cobertura temporal de las fuentes disponibles para las variables de interés. La serie homogénea elaborada en Fundación BBV (1999) cubre el período 1955-93 a intervalos bienales. Esta serie se extiende hasta 1995 en Fundación BBVA (2000), donde también se ofrecen avances de las magnitudes relevantes hasta 1999.

**Gráfico 1: Cobertura temporal de las principales series regionales**



El INE ha elaborado cuatro series diferentes de Contabilidad Regional con bases 1980, 1986, 1995 y 2000. Además del cambio en el año base, las nuevas series han incorporado cambios metodológicos para adaptarlas a las sucesivas directrices internacionales en materia de contabilidad nacional y regional, así como información adicional proveniente de encuestas nuevas o mejoradas y de otras fuentes. Las dos primera y las dos últimas series han sido enlazadas por el propio INE para construir series “homogéneas” para 1980-96 en base 1986 y 1995-2007 en base 2000.<sup>3</sup> En el primer caso, sin embargo, el enlace parece haber sido sólo parcial, pues la mayor parte de las magnitudes que se ofrecen en el sitio web del INE comienzan en 1986 (la excepción es el PIB).

En el gráfico se incluyen también dos bases de datos adicionales. La primera es la BDMores en base 2000 (de Bustos et al, 2008), donde se construye un cuidadoso enlace por ramas entre las diversas series de Contabilidad Regional. Aunque casi no utilizaré estos datos en lo que sigue, las series de la BD Mores se incluirán en los gráficos comparativos recogidos en el Anexo 2.

### ***La CRE00-r***

La última base de datos incluida en el Gráfico 1, a la que denominaré CRE00 revisada (*CRE00-r*), será la fuente de las series *terminales* de renta y empleo que extenderé hacia atrás desde 1995 hasta 1955 mediante su enlace con otras series anteriores. Esta fuente difiere de la CRE00 original únicamente en su extensión hacia atrás para los años 1995 a 1999 en los que he construido un enlace alternativo al del INE cuyos detalles se presentan en de la Fuente (2009).

<sup>3</sup> En la versión de esta serie que aquí se utiliza, los datos de 2005 y 2006 son provisionales y el de 2007 es un avance.

El problema fundamental que en mi opinión presenta el enlace del INE es que, pese a los muy considerables cambios que se producen en la estimación del VAB y el empleo regionales con el pase de la base 1995 a la nueva base 2000, las series enlazadas que construye el Instituto respetan en principio la estimación original de VAB y empleo en el año 1995 de acuerdo con la serie más antigua, excepto en lo que respecta a los efectos del cambio metodológico que se ha producido en el tratamiento de los servicios de intermediación financiera.<sup>4</sup> Esto equivale a suponer que las fuertes diferencias existentes entre la nueva serie y la anterior en el año de su enlace (2000) se han generado en su totalidad durante el quinquenio comprendido entre sus respectivos años base. Dada la magnitud de la revisión al alza de las series (especialmente en el caso del empleo), si este supuesto es incorrecto el procedimiento de enlace que ha adoptado el INE introducirá fuertes distorsiones en las tasas de crecimiento de las series enlazadas que pueden reducir considerablemente su utilidad para el análisis estadístico.

El procedimiento adoptado por el INE estaría justificado si la revisión al alza de las series se debiese únicamente a la infravaloración de la población extranjera en los años inmediatamente anteriores a la realización del Censo de 2001. Comparando las series originales de ocupados de la EPA entre 1996 y 2000 con la versión revisada de las mismas elaborada por el propio INE una vez estuvieron disponibles los resultados detallados del Censo de 2001, en el trabajo citado se concluye que este factor explica sólo una fracción de la revisión al alza del empleo en la nueva serie, lo que hace aconsejable modificar el procedimiento de enlace.

Partiendo de este resultado, en de la Fuente (2009) se construye un enlace alternativo de ambas series de la CRE. El enlace se realiza en tres etapas. En primer lugar, se corrige la serie de empleo 1995-2000 de la CRE95 en proporción a la revisión al alza de la serie de ocupados de la EPA durante estos años. En segundo lugar, la serie de VAB de la CRE en base 95 se corrige para reflejar el cambio en el tratamiento de los servicios de intermediación financiera y el resultado se ajusta al alza en proporción a la corrección realizada en la serie de empleo, utilizando la relación estimada entre las revisiones de ambas series durante el período de solapamiento de las bases 1995 y 2000. Finalmente, las series de empleo y VAB de la CRE en base 2000 se extienden hacia atrás desde su año base hasta 1995 tomando como referencia las series corregidas en base 95 y suponiendo que la discrepancia entre las dos series que aflora en el año 2000 refleja un error de medición en la serie más antigua que, en un 90%, ya existía en 1995.

---

<sup>4</sup> Los SIFMI (servicios de intermediación financiera medidos indirectamente) corresponden a los ingresos de las instituciones financieras por las comisiones de intermediación implícitas en el diferencial existente entre los tipos de interés aplicables a préstamos y depósitos y el tipo de mercado de referencia. A diferencia de lo que se hacía en versiones anteriores de la CRE (y de la CNE), en la CRE base 2000 los SIFMI se imputan a los sectores usuarios de servicios financieros, considerándose como consumo final la parte de los mismos que corresponde a los hogares, instituciones sin ánimo de lucro y administraciones públicas y como consumos intermedios la que corresponde a las empresas. Puesto que hasta el momento, los SIFMI se consideraban en su totalidad consumos intermedios de una rama ficticia y reducían el VAB agregado (pero no el sectorial), el cambio en su tratamiento se traduce en una reducción de los VABs sectoriales y en un incremento del VAB agregado. Hay que observar que el BBVA se ha anticipado al INE en lo que concierne al tratamiento de los SIFMI, habiendo adoptando desde el principio un procedimiento similar al establecido en la CRE00.

### 3. Dos cuestiones metodológicas

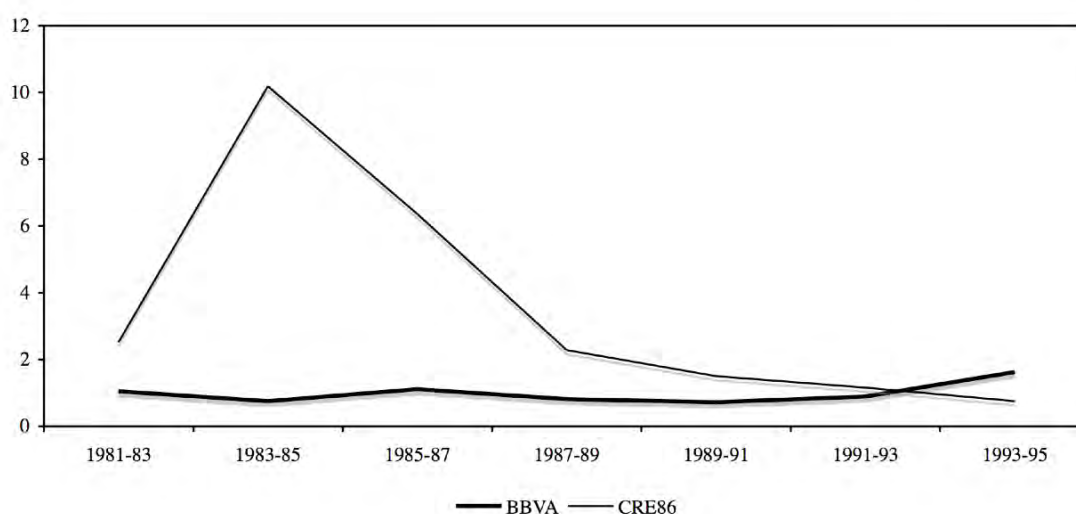
Tomando como dadas las series de la CRE00-r, el problema que se aborda en el presente trabajo es el de cómo extenderlas hacia atrás de la mejor forma posible dada la información disponible para los años anteriores a 1995. Los principales problemas que plantea el ejercicio son dos. Primero, ¿qué serie deberíamos utilizar como referencia para realizar el enlace cuando hay varias disponibles? Y segundo, ¿cómo exactamente debería construirse el enlace dado que esto siempre plantea una disyuntiva entre modificar las tasas de crecimiento o los niveles iniciales de la serie más antigua? En esta sección se describen las opciones metodológicas adoptadas en el presente trabajo en relación con ambas cuestiones.

#### 3.1. ¿Qué serie de referencia?

Como ya hemos visto, las series de la Fundación BBVA se solapan con las de Contabilidad Regional entre 1980 y 1995, lo que nos obliga a elegir entre ellas como referencia para la construcción de la serie enlazada. Dado que el INE dispone en principio de mucho mejores medios para estimar la renta y el empleo regionales que un servicio de estudios privado, en principio lo más lógico sería utilizar las distintas series de CRE disponibles desde 1980 y recurrir a las de la Fundación BBVA únicamente para el período anterior, cuando éstas son la única fuente disponible, tal como hacen Doménech et al (1999) y de Bustos et al (2008).

Un examen preliminar de los datos, sin embargo, sugiere que ésta podría no ser la mejor forma de proceder. La razón es que, tal como se muestra en el Gráfico 2, las series del INE presentan durante sus primeros años una elevada volatilidad que las hace un tanto sospechosas. Este hecho sugiere que podría ser buena idea elegir una “fecha de corte” algo más tardía para abandonar la serie del BBVA y pasar a utilizar los datos del INE como referencia para construir la serie enlazada.

**Gráfico 2: Varianza de las tasas de crecimiento anualizadas del PIB per cápita relativo**



- Nota: El PIB per cápita regional es un índice con España = 100. Las tasas de anualizadas se calculan como  $(pib_{it+2} - pib_{it})/2$  donde  $pib_{it}$  es el logaritmo del PIB per cápita relativo de la región  $i$  en el año  $t$ , normalizada por el promedio nacional. Se utilizan datos bienales porque la serie del BBVA sólo está disponible para años impares.

Una posible explicación del patrón que observamos en el Gráfico 2 es que los primeros años de la serie del INE podrían contener una elevada cantidad de ruido debido a las lógicas dificultades que comporta la puesta en marcha de una nueva serie estadística (a lo que hay que añadir en el caso del PIB la complicación que supuso la introducción del IVA en 1986). Existe, sin embargo, una explicación alternativa: las series de la CRE podrían ser más volátiles que las del BBVA no porque contengan más ruido que estas últimas, sino porque capturan mejor la volatilidad real de la economía. El perfil temporal de las volatilidades que observamos en el gráfico sugiere que no es probable que esta última sea la única explicación porque en caso contrario esperaríamos que (otras cosas iguales) la volatilidad de la serie del BBVA fuese siempre menor y no es así. Con todo, conviene no perder de vista esta posibilidad aunque sólo sea como una forma de modelizar el plus de credibilidad que en principio parece lógico atribuir a las series del INE.

En este contexto, el problema que se nos plantea se puede reformular de la manera siguiente. Sea  $y_i$  la (tasa de crecimiento de) la renta relativa por habitante en el territorio  $i$  en un período determinado. Supongamos que queremos estimar una ecuación de la forma

$$(1) \ y_i = \beta x_i + u_i$$

con datos de corte transversal de un año dado para analizar el efecto sobre el crecimiento de la renta de alguna otra variable,  $x$ . Dado que disponemos de dos estimaciones alternativas de  $y$  (las series de CRE86 y BBVA) ¿cuál de ellas deberíamos utilizar?

Para responder esta pregunta resulta útil tratar a las dos series de referencia como "proxies ruidosas" de la variable subyacente,  $y$ , en el marco de un modelo de error de medición. Para recoger las dos hipótesis indicadas más arriba, supondré que la serie de CRE sólo contiene un error de medición clásico, mientras que la serie del BBVA podría sufrir además de un sesgo sistemático hacia la media muestral (que es aproximadamente cero tal como he normalizado las variables). Tenemos, por tanto,

$$(2) \ y_i^{CRE} = y_i + e_i^{CRE}$$

$$(3) \ y_i^{BBV} = (1 - \phi)y_i + e_i^{BBV}$$

donde  $y_i$  es el valor real de la variable de interés, el parámetro  $\phi \in [0,1]$  mide el tamaño del sesgo hacia cero de las series del BBVA y  $e_i^{CRE}$  y  $e_i^{BBVA}$  son errores de medición "puros", distribuidos idéntica e independientemente con media cero y varianzas  $\sigma_{CRE}^2$  y  $\sigma_{BBV}^2$  respectivamente y no correlacionados con  $x$  e  $y$  o entre sí.

El caso más sencillo es aquel en el que las dos series contienen sólo un error clásico de medición ( $\phi = 0$ ). Bajo esta hipótesis, la serie con menor volatilidad es siempre preferible porque una mayor varianza indica necesariamente un mayor contenido de ruido y cuánto mayor sea éste, menor será la precisión con la que podemos estimar el parámetro de interés.

Cuando  $\phi > 0$  la situación se complica porque una menor varianza ya no equivale a menos ruido y porque el error sistemático en la medición de  $y$  genera un sesgo en la estimación de  $\beta$ . El

criterio natural en este caso es la minimización del error cuadrático medio (ECM) del estimador de  $\beta$ . Como veremos enseguida, bajo ciertos supuestos es posible obtener una estimación consistente del ECM del estimador de  $\beta$  que se obtendría con cada una de las series, pero únicamente si conocemos el valor de  $\phi$ , lo que no es el caso. Sin embargo, también es posible estimar (i) el sesgo mínimo que tendrían que tener los datos del BBVA para que la serie de CRE pudiese ser preferible para alguna variable  $x$ , y (ii) una cota superior para  $\phi$ . La combinación de ambas estimaciones nos da una condición suficiente pero no necesaria para establecer la superioridad de la serie del BBVA ("en valor esperado"). Aunque este resultado no elimina necesariamente toda la incertidumbre sobre qué serie es mejor, ciertamente permite acotar el problema.

Sean  $ECM_{BBV}$  y  $ECM_{CRE}$  los errores cuadráticos medios de los estimadores de  $\beta$  que se obtienen estimando por MCO la ecuación (1) con cada una de las series de referencia. Como se muestra en el Anexo 1, la función  $F()$  dada en

$$\begin{aligned}
 \frac{n-1}{\beta^2} (\widehat{ECM}_{BBV} - \widehat{ECM}_{CRE}) &= F(\phi, R^2(x)) \\
 (4) \quad &= (n-2)\phi^2 + \left( 2 + \frac{1}{R^2(x)} \left( \frac{\hat{r}_{BBV} - \hat{r}_{CRE}}{\hat{r}_{BBV}\hat{r}_{CRE}} \right) \right) \phi - \frac{1}{R^2(x)} \left( \frac{\hat{r}_{BBV} - \hat{r}_{CRE}}{\hat{r}_{BBV}\hat{r}_{CRE}} \right) \\
 &\equiv (n-2)\phi^2 + \left( 2 + \frac{A}{R^2(x)} \right) \phi - \frac{A}{R^2(x)}
 \end{aligned}$$

nos da un estimador consistente de la diferencia entre estas magnitudes, de forma que la serie del BBVA será preferible a la de la CRE si y sólo si  $F() < 0$ . En esta expresión,  $n = 18$  es el número de observaciones en el corte transversal (las 17 comunidades autónomas más Ceuta y Melilla),  $R^2(x)$  el  $R^2$  de la regresión (1) si ésta se pudiese estimar directamente con la variable  $y$  "correctamente medida" y  $\hat{r}_{BBV}$  y  $\hat{r}_{CRE}$  los valores estimados de los *ratios de fiabilidad* de las dos series, que se obtienen regresando cada una de ellas sobre la otra con datos de corte transversal para el período de interés (véase la sección 1 del Anexo 1).

Inspeccionando la ecuación (4) vemos que si  $\hat{r}_{CRE} > \hat{r}_{BBV}$  entonces  $A < 0$  y  $F() > 0$  para todo  $\phi$  y  $x$ , lo que implica que la serie del INE es siempre la mejor. Bajo el supuesto contrario (que es el que nos interesa aquí dado que es el más habitual en los datos durante el período muestral), es fácil comprobar que  $F()$  es una función creciente en  $\phi$  y en  $R^2$  en todo su dominio  $[0,1] \times [0,1]$  con

$$\begin{aligned}
 (5) \quad F(0, R^2(x)) &= -\frac{A}{R^2(x)} < 0 \\
 F(1, R^2(x)) &= n > 0
 \end{aligned}$$

tal como se ilustra en el Gráfico 3. Por tanto, si la serie del BBVA tiene un mayor ratio de fiabilidad, entonces es preferible a la del INE siempre que no presente un sesgo sistemático ( $\phi = 0$ ) y continúa siéndolo para valores suficientemente bajos de  $\phi$ . En términos generales, el valor de  $\phi$  que iguala los ECMs obtenidos con las dos series depende de  $x$ , pero existe un valor



crítico de este parámetro,  $\tilde{\phi}$  tal que la serie del BBVA es preferible para cualquier variable  $x$  siempre que  $\phi < \tilde{\phi}$ . Para calcular este valor, observamos que

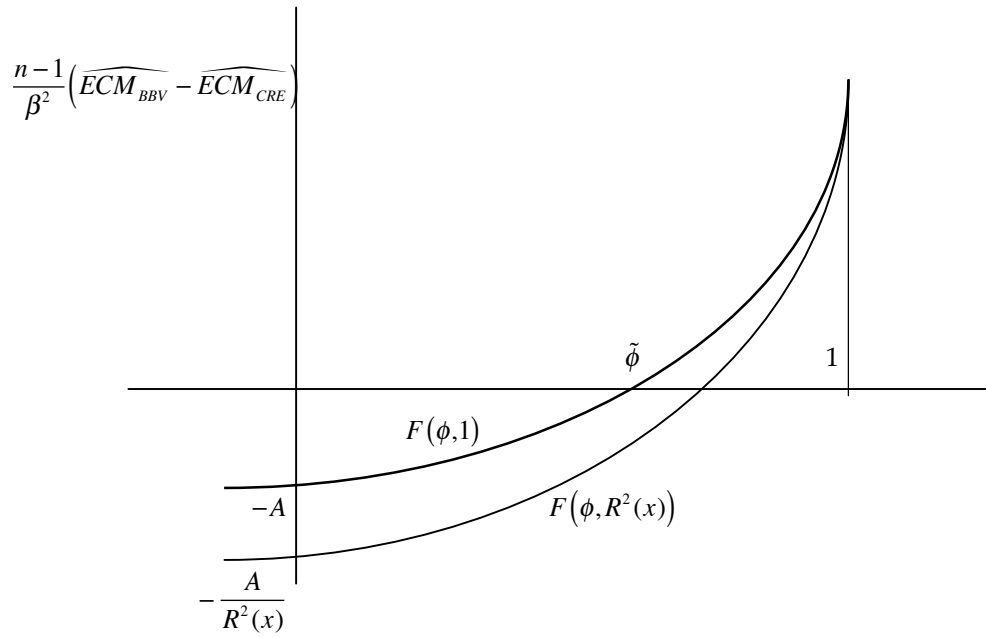
$$F(\phi, R^2(x)) \leq F(\phi, 1) = (n-2)\phi^2 + (2+A)\phi - A$$

de forma que si  $F(\phi, 1)$  es negativo también lo será  $F(\phi, R^2(x))$  para cualquier  $x$ , lo que asegura la superioridad de la serie del BBVA para cualquier regresor. El valor crítico de  $\phi$  es, por tanto, el mayor de los ceros de la función cuadrática  $F(\phi, 1)$  y viene dado por

$$(6) \quad \tilde{\phi} = \frac{-(2+A) + \sqrt{(2+A)^2 + 4(n-2)A}}{2(n-2)}$$

Alternativamente, podemos fijar un valor "razonable" de  $R^2(x)$  para obtener una estimación de  $\tilde{\phi}$  bajo un criterio algo menos exigente.

**Gráfico 3: La función  $F()$**



Por otra parte, en el Anexo 1 se muestra también que es posible estimar una cota superior para  $\phi$  utilizando

$$(7) \quad \hat{\phi}_{\max} = 1 - \max_t \hat{r}_{CRE,t}$$

donde  $\hat{r}_{CRE,t}$  es el ratio de fiabilidad estimado para la serie CRE86 en el año  $t$ . Puesto que  $\phi \leq \hat{\phi}_{\max}$ , una condición suficiente (en "valor esperado") pero no necesaria para que la serie del BBVA sea preferible a la del INE en un año dado es que  $\hat{\phi}_{\max} \leq \tilde{\phi}$ .

En lo que sigue, calcularé los valores de  $\hat{\phi}_{\max}$  y de  $\tilde{\phi}$  en cada año para distintos agregados regionales (de renta y empleo) y los utilizaré de guía para intentar determinar una fecha razonable de corte. Como cabría esperar a la vista del Gráfico 2, el procedimiento sugiere que no es aconsejable utilizar los primeros años de la serie del INE. En consecuencia, el cambio en las series de referencia se pospondrá hasta finales de los años ochenta.

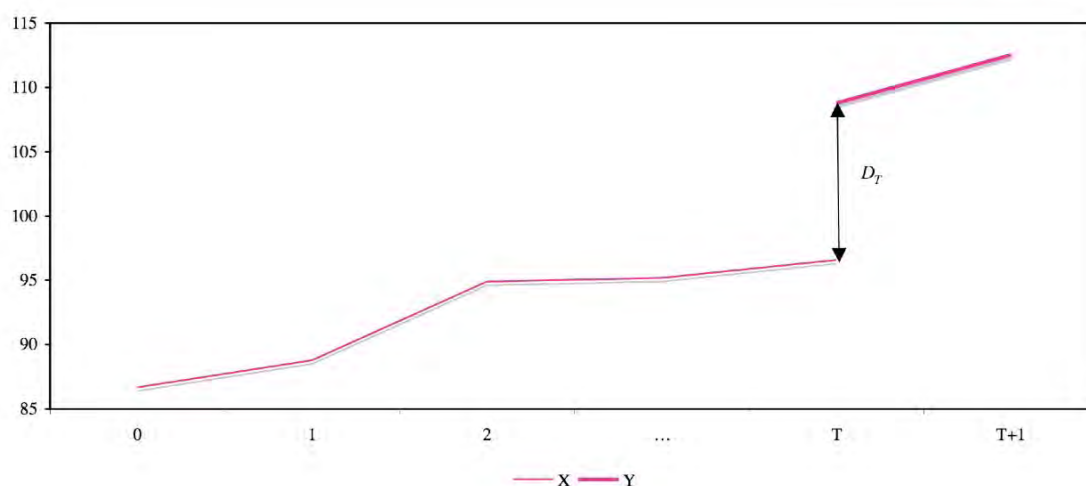
### 3.2. ¿Qué procedimiento de enlace?

Consideremos ahora los detalles de la construcción del enlace entre dos series,  $X_t$  e  $Y_t$ . La serie de construcción más reciente y en principio más fiable,  $Y_t$ , comienza en  $T$ , mientras que la serie más antigua,  $X_t$ , comienza en 0 y se extiende al menos hasta  $T$ . En el punto de enlace,  $T$ , los valores de ambas series generalmente no coincidirán debido a la introducción de cambios metodológicos o a la incorporación de nueva información de base en la construcción de la nueva serie. En el primer caso, lo más razonable es reconstruir en la medida de lo posible la serie más antigua utilizando la nueva metodología antes de realizar el enlace. En lo que sigue, supondré que esto ya se ha hecho, de forma que la diferencia existente entre las dos series en el momento de su enlace se debe únicamente a una mejora de la base estadística que permite medir mejor algunas actividades o sectores que previamente no estaban cubiertos de forma adecuada. Típicamente, esto se traduce en el "afloramiento" de nueva actividad, tal como sucede en el Gráfico 4, donde la nueva serie se sitúa claramente por encima de la más antigua en el periodo  $T$ . Llamaré  $D_T$  a la discrepancia existente entre la serie nueva y la serie antigua en el momento del enlace, y

$$(8) d_T = \ln Y_T - \ln X_T = y_T - x_T$$

a la diferencia proporcional o logarítmica entre ambas en el mismo momento.

Gráfico 4



El enlace entre las dos series -- esto es, la extensión hacia atrás de la más reciente tomando como referencia la más antigua-- se puede realizar de varias formas que responden básicamente a diversas hipótesis sobre el perfil temporal del "error de medición" contenido en la serie más

La retropolación consiste en extender hacia atrás la serie nueva utilizando las tasas de crecimiento de la serie antigua. Gráficamente, la idea es “elevar” la serie antigua de manera proporcionalmente uniforme, respetando su perfil temporal, hasta que coincida con la serie nueva en el momento del enlace, tal como se observa en el Gráfico 5. Utilizando minúsculas para indicar que estamos trabajando con logaritmos, la retropolación de  $Y_t$  tomando como referencia  $X_t$  vendría dada por

de forma que la extensión de la serie enlaza con  $Y_t$  en el año  $T$  y conserva la tasa de crecimiento de la serie  $X_t$  para los años anteriores al punto de enlace, esto es,

La hipótesis implícita en este caso es que el “error de medición” en la serie antigua,

es constante, esto es, que ya existía en el momento 0 y que su magnitud, medida en términos proporcionales, no ha variado con el paso del tiempo. Por lo tanto, para recuperar el valor “correcto” de la magnitud de interés, basta con añadir a la serie antigua (medida en logaritmos) la discrepancia proporcional observada en el momento del enlace,  $d_T$ .

11

un ritmo uniforme. En consecuencia, tendremos que sumarle a la serie antigua (en logaritmos) una fracción creciente de  $d_T$ :

$$(12) \hat{y}_t^i = x_t + \frac{t}{T}(y_T - x_T) = x_t + \frac{t}{T} d_T$$

De esta forma se conserva el valor original de la serie antigua en su año base, pero no así su tasa de crecimiento, a la que se añade una fracción constante de  $d_T$  que depende de la longitud del período transcurrido entre el año base y el momento del enlace

$$(13) \Delta \hat{y}_t^i = \Delta x_t + \frac{1}{T} d_T \text{ para } t \leq T \text{ y } \hat{y}_o^i = x_o,$$

lo que puede generar una serie enlazada con un perfil temporal muy distinto al de la serie original. (Véase el Gráfico 5).

Conviene observar que los dos procedimientos siguen un esquema común. En ambos casos, la serie enlazada se construye añadiendo a la serie más antigua una estimación del error de medición contenido en la misma que se basa en la única observación directa de esta magnitud de la que disponemos, la correspondiente al momento del enlace. Así pues, la serie enlazada por el método  $j$  vendrá dada por

$$(14) \hat{y}_t^j = x_t + \hat{d}_t^j \text{ para } t \leq T \text{ con } \hat{d}_t^j = \begin{cases} \hat{d}_t^r = d_T \\ \hat{d}_t^i = \frac{t}{T} d_T \end{cases}$$

### *¿Qué procedimiento es mejor?*

¿Cuándo deberíamos utilizar cada uno de los dos procedimientos de enlace descritos en el apartado anterior? Para justificar la elección del método de interpolación en un contexto determinado, el INE (2009d) observa que las estimaciones de la CNE y de la CRE correspondientes al año base de cada serie tienen "carácter estructural," esto es, están basadas en análisis más exhaustivos que los que se realizan en otros años y deberían, por tanto, considerarse más fiables que otras observaciones. Tal como sostiene el Instituto, este hecho aconsejaría la utilización del método de interpolación entre años base pero, en mi opinión, únicamente cuando el cambio de base no supone ninguna modificación de la metodología de estimación o de la información primaria utilizada en la construcción de las series. En la mayoría de las ocasiones, sin embargo, los cambios de base de la CNE y de la CRE se han aprovechado precisamente para introducir mejoras de ambos tipos que deberían redundar en una mayor precisión de las estimaciones. En estas circunstancias, las discrepancias observadas entre la serie de nueva elaboración y su antecesora en el año del enlace reflejarán, al menos en parte, la reducción en el error de medición de la segunda de ellas que se produce como resultado de la introducción de mejoras metodológicas y estadísticas. Parece razonable pensar que el error que aflora en el nuevo año base ya estaba presente en alguna medida en todos los años de la serie antigua y no sólo en el último de ellos. En consecuencia, la opción de preservar el valor original

de la serie más antigua en su año base, propia del método de interpolación, sólo parece aconsejable en circunstancias que generalmente no se dan en el contexto que nos interesa.

El segundo procedimiento de enlace que hemos repasado, el método de retropolación, se basa precisamente en la hipótesis de que el error aflorado en el nuevo año base ya existía desde el comienzo de la serie más antigua y ha permanecido constante en el tiempo en términos relativos. Puestos a escoger entre hipótesis extremas, ésta parece en principio más plausible que la anterior, pero sigue siendo bastante arriesgada. Aunque encontrar evidencia sobre el tema resulta muy difícil por motivos obvios, la intuición sugiere que los errores de medición de las series de contabilidad nacional y regional deberían aumentar con el tiempo porque los problemas de cobertura tienden a ser especialmente severos en sectores emergentes con un peso creciente en el agregado.

Todo esto sugiere que sería aconsejable introducir una metodología alternativa de enlace que proporcione una vía intermedia entre los dos procedimientos habituales. En el apartado siguiente se desarrolla un procedimiento con estas características, al que denominaré *método mixto* de enlace.

#### *Un procedimiento "mixto" de enlace*

Como hemos visto, la diferencia entre los dos procedimientos resumidos en la primera parte de esta sección puede expresarse en términos de las hipótesis implícitas en cada uno de ellos sobre el valor del error de medición contenido en la serie más antigua en el momento 0. En la interpolación se supone que tal error es nulo ( $d_o = 0$ ), mientras que en la retropolación la hipótesis es que este error es igual al observado en el momento del enlace ( $d_o = d_T$ ). Una opción más flexible consiste en parameterizar el error de medición inicial de una forma que permita acomodar cualquier hipótesis intermedia. Tendremos ahora

$$(15) \hat{d}_o^m = \rho d_T$$

donde  $\rho$  es un parámetro, cuyo valor podemos fijar libremente, que mide la magnitud del error inicial. Obsérvese que este procedimiento equivale a fijar el valor inicial de la serie enlazada,  $\hat{y}_o^m$ , como una media ponderada de los valores iniciales de la serie más antigua y de la serie enlazada obtenida por retropolación,

$$(16) \hat{y}_o^m = x_o + \hat{d}_o^m = x_o + \rho d_T = (1 + \rho - \rho)x_o + \rho d_T = (1 - \rho)x_o + \rho(x_o + d_T) = (1 - \rho)x_o + \rho \hat{y}_o^r$$

La segunda modificación que introduciré en relación con el procedimiento de interpolación expuesto más arriba tiene que ver con el perfil temporal del error de medición de la serie más antigua. En la interpolación se supone que este error aumenta linealmente con el tiempo. Sin embargo, esto no parece lo más plausible, al menos si la fuente del error es, como parece probable, la deficiente medición de ciertas actividades cuyo peso relativo aumenta con el paso del tiempo. En este caso, el crecimiento del error dependerá del ritmo relativo de crecimiento de tales actividades en relación con el resto de la economía. Si suponemos que la razón entre

ambas tasas de crecimiento es aproximadamente constante, podemos modelizar la evolución del error de una forma sencilla.

Supongamos que el error de medición de la serie antigua,  $D_t$ , es una fracción constante  $\theta$  del volumen de alguna actividad deficientemente medida,  $Z_t$ , de forma que el valor “real” de la serie vendría dado por

$$(17) Y_t = X_t + D_t = X_t + \theta Z_t$$

Supongamos también que la tasa de crecimiento bruto de  $Z_t$  guarda una relación constante con la tasa de crecimiento de  $X_t$ . Esto es, llamando  $G_t$  a la tasa de crecimiento bruto de  $X_t$ ,

$$(18) \frac{X_{t+1}}{X_t} = G_t$$

supondremos que

$$(19) \frac{Z_{t+1}}{Z_t} = \mu G_t$$

donde  $\mu$  es una constante cuyo valor determinaremos más adelante. Llamando  $z$  al ratio  $Z/X$ , obtenemos una sencilla ecuación en diferencias

$$(20) z_{t+1} = \frac{Z_{t+1}}{X_{t+1}} = \frac{\mu G_t Z_t}{G_t X_t} = \mu \frac{Z_t}{X_t} = \mu z_t$$

cuya solución es de la forma

$$(21) z_t = z_o \mu^t$$

Dividiendo ambos lados de (17) por  $X_t$ , tenemos

$$(22) \frac{Y_t}{X_t} = \frac{X_t + \theta Z_t}{X_t} = 1 + \theta z_t$$

Tomando logaritmos de esta expresión, llegamos a una conveniente aproximación para  $d_t$ .

$$(23) d_t = \ln \frac{Y_t}{X_t} = \ln(1 + \theta z_t) \cong \theta z_t$$

Utilizando (21), esta expresión implica que

$$(24) d_t \cong \theta z_t = \theta z_o \mu^t = d_o \mu^t$$

Finalmente, podemos calcular el valor de  $\mu$  implícito en  $\rho$ . Evaluando (24) en el momento  $T$  y recordando que, por hipótesis,  $d_o = \rho d_T$  tendremos

$$(25) d_T \cong d_o \mu^T = \rho d_T \mu^T$$

de donde

$$(26) 1 \equiv \rho \mu^T \Rightarrow \mu = \left( \frac{1}{\rho} \right)^{\frac{1}{T}}$$

Finalmente, sustituimos (26) en (24) para obtener la siguiente expresión para la senda del error proporcional de medición:

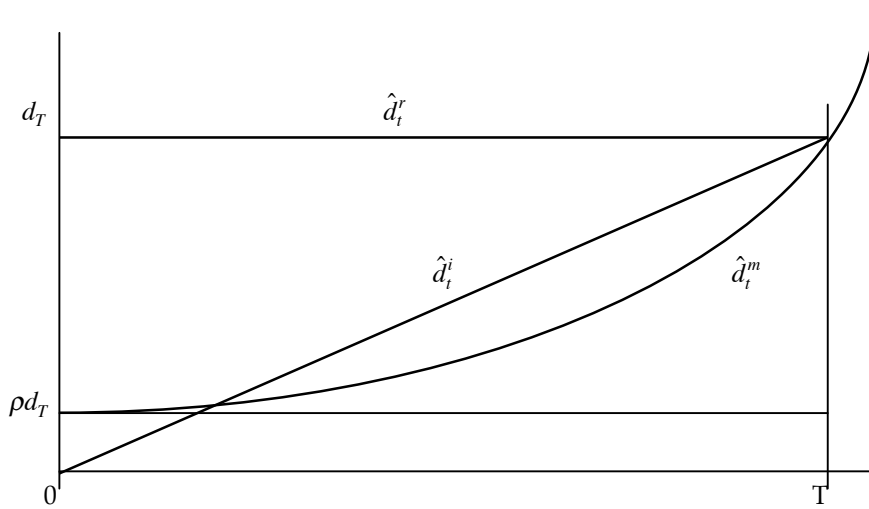
$$(27) d_t \equiv d_o \left( \frac{1}{\rho} \right)^{\frac{t}{T}} = \rho d_T \left( \frac{1}{\rho} \right)^{\frac{t}{T}} = d_T \rho^{\frac{T-t}{T}}$$

Por tanto, la serie enlazada por el procedimiento mixto vendrá dada por

$$(28) \hat{y}_t^m = x_t + \hat{d}_t^m \text{ para } t \leq T \text{ con } \hat{d}_t^m = d_T \rho^{\frac{T-t}{T}}$$

Obsérvese que la ecuación (27) describe una trayectoria del error estimado de medición que ocupa una posición intermedia entre las que se derivan de las ecuaciones análogas para los dos métodos de enlace más habituales, dadas en la expresión (14). (Véase el Gráfico 6) Bajo nuestras hipótesis en este apartado, el error ( $d_t$ ) no será constante ni lineal en  $t$  sino que aumentará a una tasa creciente con el paso del tiempo porque el peso de las actividades mal cubiertas por las estadísticas de base será mayor cuánto más cerca estemos del punto de enlace. Por lo tanto, la corrección apropiada a la tasa de crecimiento de la serie original no será constante sino creciente en el tiempo y aumentará con el valor del parámetro  $\rho$  que refleja nuestra hipótesis sobre la severidad del error inicial de medición.

**Gráfico 6: Trayectoria del error de medición estimado de la serie antigua en los distintos procedimientos de enlace**



El problema obvio que se plantea con el método mixto es el de cómo fijar el valor de  $\rho$ . Siempre que resulte posible, intentaré basar esta decisión en algún tipo de información externa que nos pueda dar alguna idea de la fiabilidad relativa de las dos estimaciones que implícitamente se utilizan en el método mixto para fijar el valor inicial de la serie enlazada (el valor original de la serie más antigua y la estimación obtenida por retropolación).

### *Enlaces de series de precios y magnitudes "reales"*

Cuando  $X$  e  $Y$  son magnitudes monetarias, se plantea también la cuestión de cómo enlazar sus valores "reales" (en volumen o a precios constantes del año base). En principio, esto puede hacerse de dos formas alternativas: enlazando los correspondientes índices de precios para deflactar después la serie enlazada a precios corrientes, o enlazando directamente las magnitudes reales para recuperar seguidamente un índice de precios para la serie enlazada.

En el caso de las series de precios, el procedimiento natural de enlace es, en la mayor parte de los casos, la retropolación. La razón es que la única información real que contienen las series de precios regionales que existen para España es la que recogen sus variaciones a lo largo del tiempo. Su nivel, sin embargo, es totalmente arbitrario y generalmente se fija igualando a la unidad para todas las regiones el valor del índice en el año que se toma como base. Las discrepancias de nivel entre series de precios en el momento del enlace, por tanto, tampoco contienen información alguna porque reflejan únicamente el hecho de que las series a enlazar tienen distintos años base. Si eliminamos esta diferencia, tomando como referencia común para ambas series el año del enlace, la diferencia entre ellas en ese año será siempre cero por definición. No tiene mucho sentido, por tanto, pensar en cómo debemos distribuir el "error" aflorado en el momento del enlace a lo largo del tiempo.<sup>5</sup>

En consecuencia, la forma más sencilla de construir series enlazadas de magnitudes reales consiste en enlazar en primer lugar las magnitudes medidas a precios corrientes por el procedimiento que se considere más oportuno para, seguidamente, enlazar por retropolación los correspondientes índices de precios y, finalmente, deflactar la serie nominal construida en la primera etapa utilizando esta serie de precios.

Alternativamente, podemos enlazar directamente las series a precios constantes (o las medidas de volumen expresadas en precios de un año de referencia) y recuperar la serie enlazada de deflactores a posteriori como el cociente entre la serie enlazada nominal y la correspondiente magnitud real. El procedimiento es algo más trabajoso porque tenemos que empezar por expresar ambas series en precios del año común (el de enlace). Los detalles se presentan en la sección 2 del Anexo 1, donde también se muestra que los valores de  $\rho$  utilizados para enlazar las series nominales y reales han de ser los mismos si queremos que la serie resultante de deflactores coincida con la que se obtendría por retropolación.

#### **4. Construcción de las series de población**

Tradicionalmente, los censos españoles han ofrecido información sobre dos agregados poblacionales diferentes: la población de hecho y la población de derecho o residente. La población de derecho de un territorio está formada por las personas que tienen su residencia habitual en el mismo. La población de hecho, por su parte, está integrada por las personas que se encuentran físicamente en dicho territorio en el momento del recuento censal. Este segundo

---

<sup>5</sup> La situación sería diferente si dispusiéramos de índices que capturasen las diferencias existentes entre regiones en niveles de precios y no sólo en sus tasas de crecimiento. En tal caso, los niveles de las series regionales en relación al deflactor agregado sí contendrían información útil y podría ser conveniente utilizar algún otro método de enlace.



agregado se obtiene restando de la población de derecho los residentes habituales que se encontraban ausentes en el momento del censo y sumándole los transeúntes que se encontraban presentes en la demarcación en la misma fecha. Desde 2001, el censo español ha dejado de cuantificar los transeúntes y ofrece únicamente cifras de población residente, sin distinguir entre residentes presentes y ausentes.

El agregado que más nos interesa a efectos de relativizar la producción y el empleo es la población residente o de derecho, que es la que en principio ejerce su actividad económica habitual dentro del territorio de referencia. Sin embargo, las estadísticas de población de derecho plantean un problema no despreciable. Tal como observa Maluquer (2008), estas cifras suelen contabilizar en su territorio de origen a una parte difícil de determinar pero potencialmente significativa de la población emigrante, que en algunos casos conserva allí su residencia administrativa pese a vivir realmente en el extranjero (o en otra región). Como resultado de este fenómeno, las cifras de población de derecho tenderán a sobreestimar la población “real” del país en su conjunto y podrían ofrecer una imagen distorsionada de su distribución regional. La magnitud del problema seguramente se ha ido reduciendo con el paso del tiempo debido a la mejora de los procesos de registro y a la gradual reducción y eventual inversión del saldo migratorio neto de nuestro país. Aún así, la distorsión puede haber sido significativa durante buena parte del período que nos ocupa.

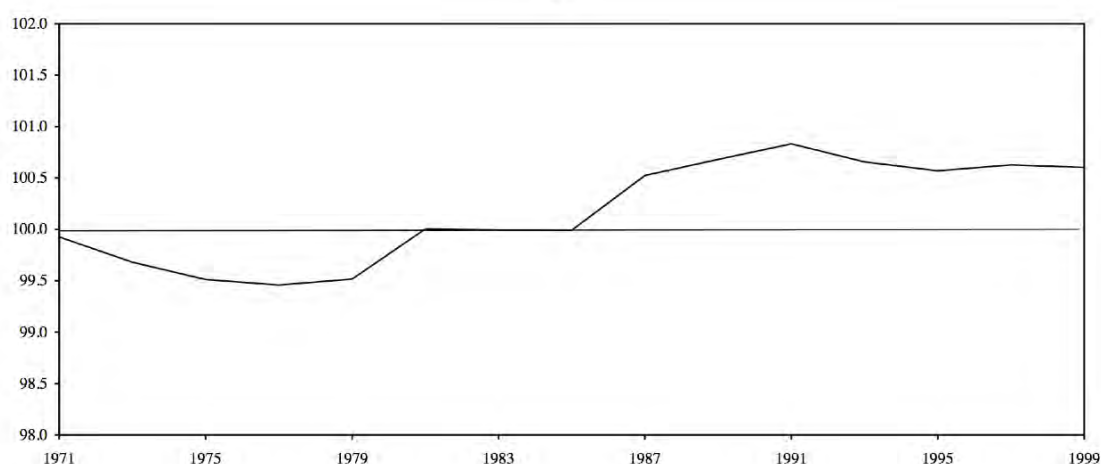
Esta circunstancia ha aconsejado la elaboración de tres series diferentes de población nacional y regional: la población de derecho, la población de hecho y lo que llamaré *población residente corregida*. Esta última serie intenta aproximar la población que realmente reside de forma habitual en cada territorio, substrayendo de la población de derecho una estimación de la emigración no registrada basada en la información disponible sobre residentes ausentes en el momento censal.

#### **4.1. Población de derecho o residente**

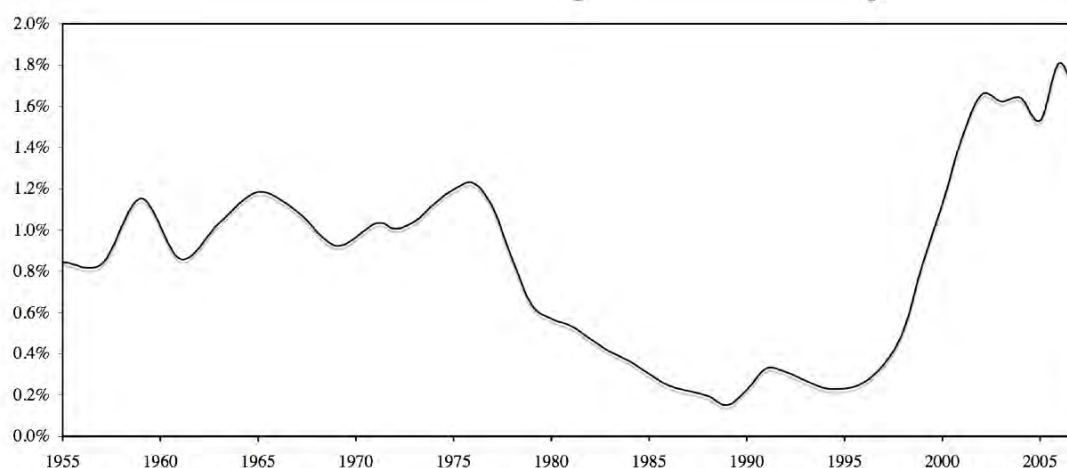
Las series de referencia del INE en materia de población de derecho son las estimaciones intercensales de población, que cubren el período 1971-2001, y las estimaciones de la población actual a partir del Censo de 2001, que extienden la serie anterior hasta el presente (véase INE, 2009c). Ambas series pueden obtenerse a 1 de julio de cada año. Para años anteriores a 1971 no se dispone (al menos en la página web del INE) de series anuales sino únicamente de datos referidos a años censales. Por otra parte, la Fundación BBVA (1999, 2000) ofrece series bienales de población de derecho a 1 julio para el período 1955-99 que, en principio, también están basadas en datos primarios del INE.

Aunque las dos fuentes son plenamente consistentes en principio, en la práctica existen diferencias apreciables entre ellas durante su período de solapamiento. Tal como se observa en el Gráfico 7, las diferencias se estrechan en años censales (con la llamativa excepción de 1991) y se amplían en otros años, lo que sugiere que su fuente principal podría ser la utilización de distintos métodos de interpolación intercensal.

**Gráfico 7: Población nacional según el BBVA (dato del INE = 100)**



**Gráfico 8: Tasa de crecimiento de la población de derecho española**



La serie de población de derecho que utilizaré en este trabajo se construye enlazando las series del INE y del BBVA. Para 1971 y años posteriores se toma directamente la serie del INE mientras que para 1961 y años anteriores se utiliza la del BBVA. Entre 1961 y 1971, las dos series se enlazan por el procedimiento de interpolación con corrección uniforme de la discrepancia final (la observada entre ambas series en 1971). Esta última es muy reducida, situándose por debajo del 0,4% en todas las regiones (con la excepción de Ceuta y Melilla, donde se eleva al 1,08%). En enlace se realiza de forma independiente para cada territorio y los resultados se suman para obtener la serie nacional de población residente. La tasa de crecimiento de esta serie se muestra en el Gráfico 8, donde se aprecia la rápida desaceleración del crecimiento de la población entre 1975 y 1995, seguida de un fuerte repunte debido al elevado influjo de inmigrantes registrado durante los últimos años del período muestral.

#### **4.2. Población de hecho**

Las series de población de hecho se construyen a partir de los datos regionales correspondientes a años censales que ofrece el INE en su página web (“series históricas de población” en INE, 2009c) y de la serie anual de población de hecho a nivel nacional que ha elaborado Maluquer

(2008). Este autor construye series anuales partiendo de los datos censales y utilizando un método de “inventario permanente” en el que se hace uso de la información existente sobre movimientos naturales de población y saldos migratorios. La discrepancia que aflora al confrontar la estimación preliminar “de flujo” construida de esta forma con cada nuevo censo se distribuye de manera uniforme sobre el período intercensal.

El primer paso en la construcción de las series regionales de población de hecho consiste en proyectar los datos de cada censo al 1 de julio del primer año impar de cada década<sup>6</sup> con el fin de disponer de una periodificación consistente con la utilizada en el caso de la población de derecho. Dado que los censos de 1950, 1960 y 1970 tienen como fecha de referencia el 31 de diciembre y los de 1981 y 1991 el 1 de marzo, necesitamos estimar el crecimiento de la población de hecho de cada región durante un período de 6 o 4 meses respectivamente para llegar al 1 de julio.

Este ajuste se realiza de la forma siguiente. En primer lugar, se calcula la tasa de crecimiento logarítmica anual de la población de *derecho* de cada región durante el período más corto posible que incluye el momento censal. Así, por ejemplo, en el caso del censo de marzo de 1981, se toman los datos de población de derecho correspondientes al 1 de julio de 1980 y de 1981 en la serie enlazada y la deseada tasa de crecimiento anual se calcula como la diferencia logarítmica entre la población en el segundo año y en el primero. Las tasas de crecimiento de la población de derecho regional obtenidas de esta forma se normalizan seguidamente por la tasa de crecimiento de la población de derecho nacional calculada con los mismos datos. A continuación, el mismo procedimiento se utiliza para calcular la tasa de crecimiento de la población *de hecho* nacional durante el año que incluye el momento censal utilizando la serie de Maluquer. Multiplicando esta tasa nacional por los índices relativos regionales, se obtiene una estimación definitiva de la tasa de crecimiento de la población de hecho de cada región que se utiliza para llevar la población censal de hecho de cada región hasta el 1 de julio más próximo al momento censal (pero siempre posterior a él). De esta forma, se preserva la tasa de crecimiento de la población nacional de hecho durante el año censal (de acuerdo con la estimación de Maluquer), utilizándose los datos de población de derecho únicamente para estimar los crecimientos poblacionales relativos de las distintas regiones.

Las cifras de población de hecho así obtenidas son directamente comparables con las de población de derecho descritas en el apartado anterior. El Cuadro 1 muestra la discrepancia porcentual entre ambos agregados al comienzo de cada década. En algunos casos (y muy especialmente en el de Galicia), la población de hecho se sitúa significativamente por debajo de la de derecho, lo que sugiere que la emigración no registrada puede haber sido importante.

Las series anuales de población de hecho regional se construyen interpolando los pesos regionales en la población de hecho española entre años censales de una forma que conserva los perfiles temporales de los pesos regionales en la población de derecho. En primer lugar, se calcula el peso de cada región en la población de hecho nacional en 1955 y en el primer año de

---

<sup>6</sup> El censo de 1950 se proyecta hasta el 1 de Julio de 1955, que es el punto de partida de las series de población de la Fundación BBVA. En este caso, la tasa relativa de crecimiento de la población de cada región se estima a partir de los datos que ofrece Alcaide (2007) para 1950 y 1955.

cada década subsiguiente utilizando los datos descritos arriba. Seguidamente, se interpola entre años “censales” de la forma que se indica a continuación. Una vez completadas las series de pesos regionales, se recupera finalmente el nivel de población multiplicando tales pesos por la población de hecho nacional de acuerdo con Maluquer.

**Cuadro 1: Discrepancia entre la población de hecho y de derecho a 1 de julio**

	1961	1971	1981	1991
<i>Andalucía</i>	-0.43%	-0.65%	0.00%	1.43%
<i>Aragón</i>	0.34%	-0.09%	1.34%	2.59%
<i>Asturias</i>	-0.20%	-0.74%	-0.11%	0.25%
<i>Baleares</i>	-0.27%	4.33%	4.77%	4.90%
<i>Canarias</i>	-1.62%	3.63%	5.91%	9.51%
<i>Cantabria</i>	-0.09%	-0.60%	-0.44%	0.49%
<i>Cast. y León</i>	-1.44%	-1.98%	-0.49%	0.57%
<i>Cast. la Mancha</i>	-1.04%	-1.90%	-1.57%	-0.53%
<i>Cataluña</i>	0.41%	0.57%	0.32%	0.87%
<i>Valencia</i>	-1.08%	-0.22%	0.17%	1.74%
<i>Extremadura</i>	-0.76%	-2.46%	-1.72%	-0.60%
<i>Galicia</i>	-3.47%	-3.79%	-1.96%	-0.55%
<i>Madrid</i>	2.44%	1.22%	1.16%	1.62%
<i>Murcia</i>	-0.13%	-0.50%	0.31%	1.23%
<i>Navarra</i>	-1.29%	-0.46%	-0.28%	0.65%
<i>País Vasco</i>	0.47%	1.08%	-0.12%	0.14%
<i>Rioja</i>	-0.39%	-0.03%	-0.44%	1.70%
<i>Ceuta y Melilla</i>	8.68%	6.55%	8.63%	9.98%
<i>España</i>	-0.47%	-0.32%	0.27%	1.37%

- Nota: (población de hecho - población de derecho)/población de derecho

El procedimiento de interpolación es el siguiente. Sea  $POBD_{it}$  la población de derecho de la región  $i$  en el año  $t$  y  $wd_{it}$  su peso en el total nacional en términos de este agregado. De la misma forma, sea  $POBH_{it}$  la población de hecho de la región y  $wh_{it}$  el peso correspondiente.  $POBD$  y  $wd$  son conocidos en todos los años de la muestra, mientras que  $POBH_{it}$  y  $wh_{it}$  se conocen sólo para los años censales (excepto por el total nacional,  $POBH_t$ ) y han de interpolarse entre ellos.

Fijemos dos años censales consecutivos, 0 y  $T$ , y consideremos un año intermedio,  $t$ . Para cada  $t$  construimos

$$(29) \Delta_{it} = \frac{wd_{it} - wd_{i0}}{wd_{iT} - wd_{i0}}$$

de forma que  $\Delta_{it}$  mide la fracción del incremento total de  $wd$  entre 0 y  $T$  que se ha producido en el momento  $t$ . El valor de  $wh$  en  $t$  se estima como

$$(30) \widehat{wh}_{it} = wh_{i0} + \Delta_{it} (wh_{iT} - wh_{i0})$$

mientras que la población de hecho total viene dada por

$$(31) \widehat{POBH}_{it} = \widehat{wh}_{it} POBH_t$$

donde  $POBH_t$  es la población de hecho nacional de acuerdo con Maluquer (2008).

#### 4.3. Población residente ajustada

La serie de población de hecho construida en el apartado anterior no proporciona una solución plenamente satisfactoria del problema derivado de la existencia de migraciones no registradas. Por un lado, puede aminorarlo al excluir de la población de derecho a los residentes ausentes (siempre que la mayoría de ellos sean realmente emigrantes no registrados), pero por otro lado introduce una distorsión adicional al incluir en la población de cada región a los transeúntes que se encuentran en ella de forma ocasional (turistas y otras personas desplazadas por motivos de trabajo o personales). En casos como el gallego, seguramente la población de hecho es un mejor indicador de la población “real”, mientras que en otros, como Baleares, casi ciertamente sucede lo contrario. Una posible solución aproximada podría ser la de utilizar en cada caso la menor de las dos series.

**Cuadro 2: Residentes ausentes como porcentaje de la población de derecho**

	1950	1960	1970	1981	1991
<i>Andalucía</i>	2.10%	2.02%	2.25%	2.07%	2.10%
<i>Aragón</i>	3.81%	2.96%	2.35%	2.29%	2.57%
<i>Asturias</i>	2.20%	1.42%	1.76%	1.69%	1.73%
<i>Baleares</i>	1.88%	1.13%	0.48%	0.79%	0.84%
<i>Canarias</i>	4.10%	5.21%	3.82%	2.41%	2.37%
<i>Cantabria</i>	2.16%	2.11%	1.69%	1.86%	1.90%
<i>Cast. y León</i>	3.38%	3.75%	3.22%	3.48%	4.25%
<i>Cast. la Mancha</i>	3.00%	2.81%	2.41%	2.91%	3.19%
<i>Cataluña</i>	1.32%	1.10%	0.99%	0.91%	1.11%
<i>Valencia</i>	1.43%	1.80%	1.32%	1.31%	1.30%
<i>Extremadura</i>	2.50%	3.57%	3.63%	3.76%	3.93%
<i>Galicia</i>	4.93%	5.68%	4.77%	3.65%	3.07%
<i>Madrid</i>	1.28%	1.64%	1.28%	1.36%	0.89%
<i>Murcia</i>	0.89%	1.28%	1.29%	1.49%	1.54%
<i>Navarra</i>	4.97%	4.79%	2.38%	3.04%	3.29%
<i>País Vasco</i>	2.23%	1.68%	1.28%	1.59%	1.86%
<i>Rioja</i>	2.60%	2.66%	1.36%	2.67%	2.13%
<i>Ceuta y Melilla</i>	1.90%	2.57%	3.06%	3.49%	3.93%
<i>España</i>	2.53%	2.58%	2.13%	1.99%	2.00%

-Fuente: INE (2009c) y Censo de 1981.

Otra posibilidad es la de corregir la población de derecho por los residentes ausentes. El Cuadro 2 muestra la información sobre esta magnitud que he podido extraer de los censos.<sup>7</sup> Comparando este cuadro con el anterior, se aprecia con claridad la distorsión que generan los

<sup>7</sup> Los censos de 1950, 1960, 1970 y 1981 contienen un desglose de la población de cada provincia en residentes presentes, residentes ausentes y transeúntes. Para 1991 no he encontrado un cuadro similar en la web del INE. Sin embargo, el INE (2009c) ofrece para ese año datos de población de derecho, transeúntes y población de hecho (este último tomado de las “series históricas de población” y no directamente de la sección sobre el censo de 1991 de la página web del Instituto) de los que se puede recuperar el número de residentes ausentes utilizando la siguiente relación:

$$pob. de hecho = pob. de derecho - residentes ausentes + transeúntes$$

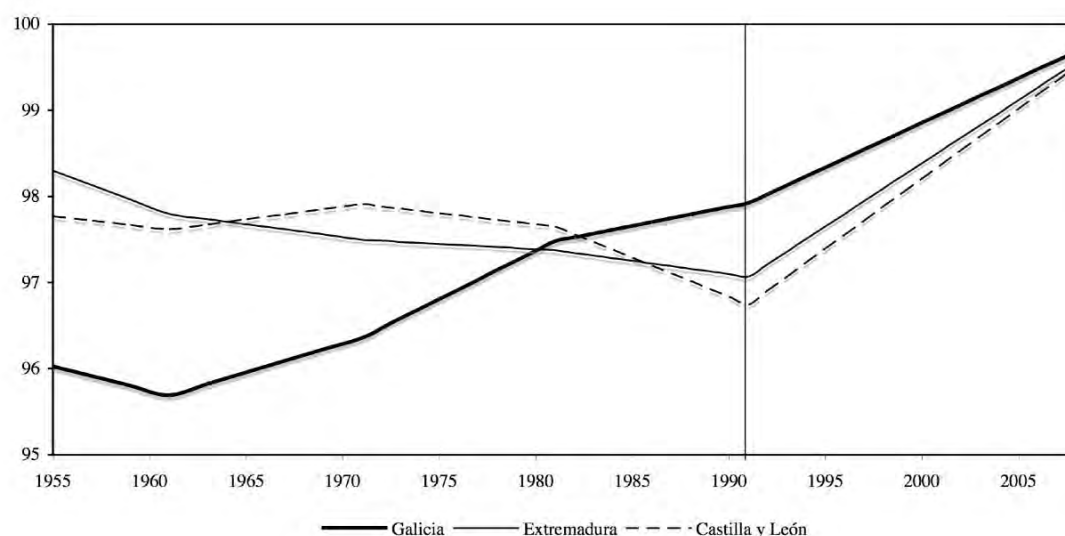
transeúntes en las regiones más turísticas. En Canarias, por ejemplo, el peso de los residentes ausentes es importante. Sin embargo, el influjo de turistas lo es todavía más, de forma que la población de hecho de Canarias se sitúa muy por encima de su población de derecho a pesar del elevado número de residentes ausentes que existe en el archipiélago

Los datos recogidos en el Cuadro 2 pueden utilizarse para introducir una corrección aproximada y parcial en las cifras de población de derecho que debería acercarnos a la población residente “real” de cada región. La corrección será sólo parcial porque al menos una parte de los presuntos emigrantes que estamos sustrayendo de la población de ciertas regiones debería añadirse a la población de las comunidades receptoras. Sin embargo, la corrección necesaria no se ha podido realizar porque no he encontrado una forma razonable de estimar cuántos de ellos residen en territorio nacional y dónde lo hacen.

No todos los residentes ausentes de su demarcación en el momento del censo son emigrantes no registrados. Muchas de las ausencias pueden reflejar desplazamientos temporales por motivos de ocio, trabajo u otras razones personales. Para intentar aproximar tales desplazamientos, he tomado como referencia el promedio de las tasas de ausencia en la fecha censal de Cataluña y Madrid. Sin contar a Baleares (a la que excluyo porque su carácter insular puede convertirla en una comunidad atípica desde el punto de vista de la movilidad de su población residente), se trata de las dos comunidades con la tasa de ausencia más reducida en promedio, tal como cabría esperar tratándose de las principales receptoras netas de población.

Restando de las cifras recogidas en el Cuadro 2 la referencia citada, se obtiene una estimación de la fracción de la población de derecho de cada región que realmente no reside en ella de forma habitual. Cuando el resultado de la operación es negativo, así como en los casos de Cataluña y Madrid, este porcentaje se iguala a cero y el dato original de población residente no se corrige. En los demás casos, la serie original de población de derecho se reduce en la proporción correspondiente. La serie de porcentajes de ausencia corregidos se completa interpolando linealmente entre años censales.

**Gráfico 9: Efectos del ajuste a la población de derecho. (Serie corregida/serie original)\*100**



- Nota: Los datos censales sobre residentes ausentes terminan en 1991.

Puesto que el censo de 2001 no ofrece información alguna sobre residentes ausentes, en principio la serie no puede extenderse más allá de 1991. A efectos ilustrativos, y con el fin de disponer de una serie “completa” que no presente saltos demasiado bruscos, he supuesto que la tasa de ausencia corregida converge a cero en 2011 para todas las regiones y he interpolado linealmente entre 1991 y este año. El Gráfico 9 ilustra los efectos de la corrección en algunas de las regiones donde las tasas de ausencia son más elevadas.

## 5. Series agregadas de empleo y VAB

En esta sección se construyen series enlazadas de empleo (puestos de trabajo) y VAB a precios básicos/coste de los factores medido a precios corrientes y constantes para el conjunto de España excluyendo a la “extra-regio.”<sup>8</sup> He decidido trabajar con el VAB a coste de los factores/precios básicos en vez de con el PIB porque el primero es el agregado de renta menos afectado por los cambios metodológicos que ha experimentado la Contabilidad Regional y, puesto que excluye el IVA, también por las modificaciones de nuestro sistema fiscal.<sup>9</sup>

La metodología utilizada es la misma en ambos casos. En primer lugar, se construye una *serie de referencia* para el período 1955-95 extendiendo hacia atrás las series de la Contabilidad Nacional de España en base 1986 (CNE86) con ayuda de algunos estudios en los que se elaboran series históricas de las magnitudes de interés mediante el enlace de las sucesivas series de la propia CNE. Seguidamente, la serie de referencia se enlaza con la correspondiente *serie terminal* tomada de la CRE00-r utilizando el método mixto descrito en la sección 3.2. En este último paso, las estimaciones nacionales de la Fundación BBVA para 1955 se utilizan como referencia para fijar el valor del parámetro  $\rho$  junto con otros datos relevantes.

Las series de VAB a coste de los factores (VABcf) a precios corrientes y constantes de 1986 de la CNE86 se extienden hacia atrás hasta 1955 utilizando las series históricas de distintos agregados de producción elaboradas por Uriel, Moltó y Cucarella (2000), mientras que la serie de empleo

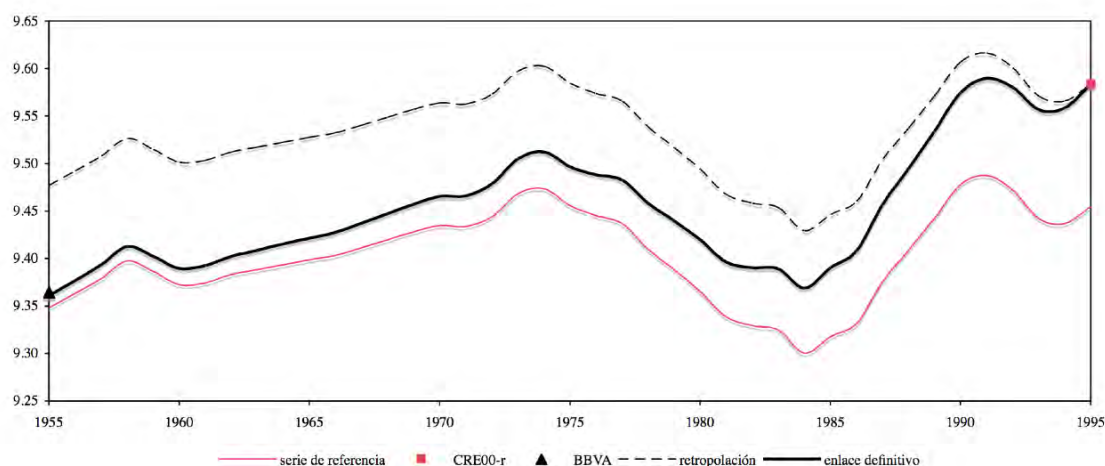
---

<sup>8</sup> La “extra-regio” o territorio extrarregional es un territorio ficticio al que se imputan aquellas actividades económicas que no se pueden asignar a una región determinada. En ella se incluyen, por ejemplo, las embajadas españolas y las bases militares y científicas en países extranjeros.

<sup>9</sup> El Valor Añadido Bruto se define genéricamente como la diferencia entre el valor de la producción y el coste de adquisición de los consumos intermedios. Dependiendo de cómo se valore la producción, se obtienen distintos conceptos de valor añadido que difieren entre sí en el tratamiento dado a ciertos impuestos indirectos y a las subvenciones a las empresas. Antes de la introducción de la base 1995, los dos principales agregados de valor añadido ofrecidos en la Contabilidad Nacional y Regional eran el VAB a precios de mercado (VABpm) y el VAB a coste de los factores (VABcf). El primero de estos agregados incluía los impuestos sobre la producción (aunque no el IVA ni los impuestos sobre las importaciones) y excluía las subvenciones a las empresas, mientras que el segundo incluía las subvenciones (fundamentalmente las ayudas de la política agrícola comunitaria y las subvenciones de explotación a empresas públicas) y excluía los impuestos sobre la producción (fundamentalmente los impuestos especiales y sobre transmisiones patrimoniales y actos jurídicos documentados así como otros impuestos menores). Comenzando con la base 1995, el principal agregado de valor añadido que se ofrece en la CRE y la CNE pasa a ser el VAB a precios básicos (VABpb). Esta magnitud es similar al VAB a coste de los factores pero no idéntica. En particular, el VABpb excluye los principales impuestos sobre los productos pero incluye algunos impuestos menores ligados a la producción (como los impuestos sobre actividades económicas y sobre bienes inmuebles). De la misma forma, se incluyen las principales subvenciones a las empresas, pero se excluyen algunas partidas menores (como las ayudas al fomento del empleo y las bonificaciones de intereses). Dado el reducido importe de estas partidas, el VABpb y el VABcf son los indicadores de valor añadido más comparables entre sí en las series de CRE anteriores y posteriores a la introducción de la base 95 (INE, 2009d).

(ocupados) de la misma fuente se completa con la serie que ofrecen Maluquer y Llonch (2005). En las series de referencia resultantes se introducen dos correcciones que intentan hacerlas más homogéneas con las correspondientes series terminales. La primera corrección consiste en substraer del VAB y del empleo una estimación de la parte de cada magnitud que corresponde a la extra-regio. La segunda intenta homogeneizar el tratamiento de los servicios de intermediación financiera en ambas series de VAB, corrigiendo al alza la más antigua en base a una estimación de la fracción de la PISB (producción imputada de servicios bancarios, antecesora de los SIFMI actuales) que se integraría en el VABcf con las nuevas normas de Contabilidad Nacional. (Véase la nota 4 y la sección 3 del Anexo 1 para más detalles).

**Gráfico 10: Enlace de la serie agregada de empleo  
escala logarítmica**



El Gráfico 10 muestra el enlace de la serie de empleo junto con la serie original de referencia, la serie que se obtendría por retropolación de la CRE00-r y los valores de esta última serie en 1995 y de la serie de empleo del BBVA en 1955. Como se aprecia en el gráfico, la diferencia entre la serie de referencia y la CRE00-r en el momento del enlace es muy significativa ( $d_T = d_{95} = 12.89\%$ ). Una parte importante de esta diferencia se debe sin duda al cambio en el concepto de empleo que se produce en ese momento: mientras que la CNE86 ofrece información sobre el número de personas ocupadas, la CRE00-r lo hace sobre el número de puestos de trabajo.<sup>10</sup> Puesto que una misma persona puede tener varios puestos de trabajo, la primera de estas magnitudes es necesariamente inferior a la segunda. De acuerdo con los técnicos del INE, la diferencia entre ambas magnitudes se sitúa actualmente en torno al 5 o 6%.<sup>11</sup> Por otro lado, la serie del BBVA, que en principio recoge el número de puestos de trabajo, se sitúa sólo un 1.63% por encima de la serie de la CNE86, lo que sugiere que el error en la serie de referencia puede

<sup>10</sup> Antes de la introducción de la base 95, la serie de empleo de la CRE y la CNE corresponde básicamente al número medio de personas ocupadas durante el conjunto del año. La correspondencia, sin embargo, no es exacta. De acuerdo con Cañada (1995), en aquellos sectores en los que la estimación de empleo de la CNE se basa en la EPA, se utilizan efectivamente los ocupados. Sin embargo, en algunos sectores en los que se utilizan otras fuentes alternativas, se parte del número de puestos de trabajo en el sector. El agregado, por tanto, mezcla ambos conceptos (ocupados y puestos de trabajo), aunque parece estar más cerca del primero. Comenzando con la base 95, sin embargo, la variable ofrecida pasa a ser el número de puestos de trabajo, con lo que una vez más el INE termina acercándose a la metodología del BBVA.

<sup>11</sup> En promedio entre 2005 y 2007, la diferencia entre la cifra de ocupados de la EPA (en media anual) y el dato de puestos de trabajo de la CRE00 era del 6.04% para España en su conjunto.



haber sido muy reducido al comienzo del período muestral y aconseja fijar un valor reducido de  $\rho$ . El censo de 1950 proporciona otro dato que apunta en la misma dirección: de acuerdo con esta fuente, sólo el 0.55% de la población activa declaraba ejercer una “actividad secundaria” además de su actividad principal, lo que sugiere un valor de  $\rho$  de en torno a 0.10 para este componente del error de medición de la serie más antigua.

El resto de la discrepancia existente entre las estimaciones de empleo de la CNE86 y de la CRE00-r en 1995 parece deberse fundamentalmente a la afloración de un volumen importante de empleo a tiempo parcial (en gran parte femenino) que comienza a detectarse gracias a una revisión del cuestionario de la EPA introducida hace pocos años para intentar captar mejor tales situaciones.<sup>12</sup> Puesto que la tasa de participación femenina ha aumentado muy significativamente con el tiempo, y seguramente lo mismo ha sucedido con la utilización de contratos a tiempo parcial, el error de estimación de la serie original que está relacionado con este fenómeno es casi ciertamente creciente en el tiempo y podría haber sido muy reducido al comienzo del período muestral. Por lo tanto, un valor de  $\rho$  de 0.10 podría ser razonable también para este segundo componente de la discrepancia entre ambas series. En consecuencia, el enlace de la serie de empleo se realiza con este valor del parámetro  $\rho$ .

En el caso del VAB, la discrepancia entre las dos series a enlazar es bastante más reducida y, dado el perfil fuertemente creciente de las series, resultaría prácticamente inapreciable en una figura análoga al Gráfico 10.

**Cuadro 3: Discrepancias entre las distintas series y valor asignado a  $\rho$**

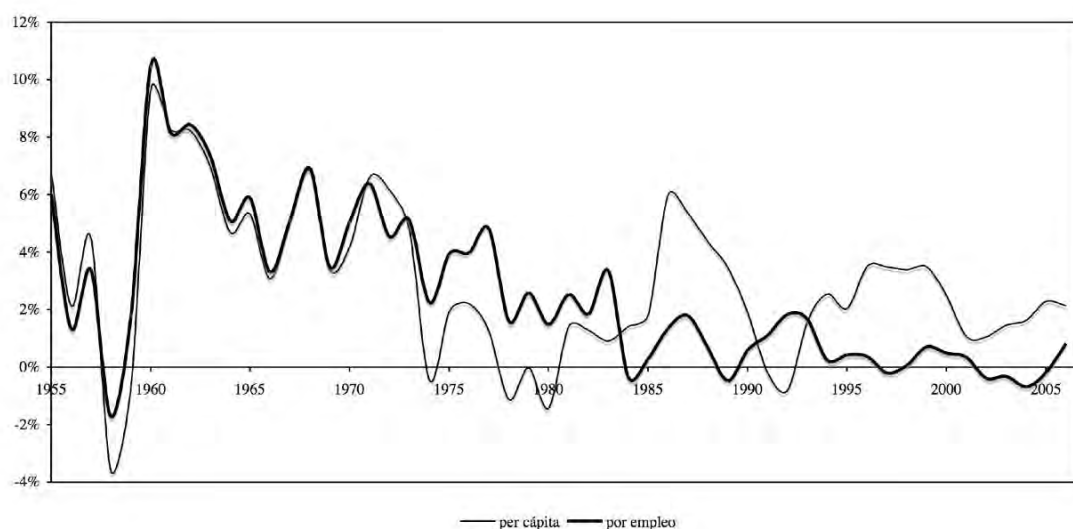
	<i>Empleo</i>	<i>VAB a precios corrientes</i>	<i>VAB a precios de 1995</i>
$d_{95} = \ln CRE00r_{95} - \ln CNE86_{95}$	12.89%	4.90%	4.90%
$d_{55}^{BBVA} = \ln BBVA_{55} - \ln CNE86_{55}$	1.63%	8.09%	11.06%
<i>valor asignado a <math>\rho</math></i>	0.10	0.90	0.90

El Cuadro 3 resume las discrepancias en años seleccionados entre las distintas series con las que estamos trabajando y muestra el valor asignado a  $\rho$  en cada caso. La primera fila muestra la diferencia logarítmica entre la serie terminal (CRE00-r) y la serie de referencia (CNE86) en el momento del enlace, que es también la diferencia porcentual entre la serie retropolada y la serie de referencia en 1955. La segunda fila muestra la diferencia logarítmica entre la estimación del BBVA y la serie de referencia en 1955. Antes de operar con ellas, las series de VAB a precios constantes se expresan en precios de 1995, que es el año común entre ambas series. En 1955 las series de VAB del BBVA se sitúan por encima no sólo de la estimación de la CNE86 sino también de la serie retropolada, lo que refuerza la hipótesis de que la infravaloración del VAB no es un fenómeno de origen reciente. En consecuencia, he elegido un valor de  $\rho$  cercano a la unidad (0.90), con lo que el enlace realizado se aproxima mucho a una retropolación pura.

<sup>12</sup> La discrepancia entre la serie corregida de ocupados de la CNE86 y la serie original de puestos de trabajo de la CNE95 en 1995 es del 7.2%. El resto de la discrepancia, hasta el 13%, se debe a la revisión al alza del empleo de la CNE95 al enlazarla con la CNE00, donde ya se recogen los efectos de la revisión de la EPA mencionada en el texto.

Utilizo un mismo valor de  $\rho$  para las series de VAB a precios corrientes y constantes (que tienen el mismo  $d_{95}$  por estar ambas medidas en las mismas unidades en el punto de enlace) porque, como hemos visto, esto equivale a construir las series enlazadas de deflatores por retropolación, que es en principio el método más apropiado en este caso.

**Gráfico 11: Tasa de crecimiento del VAB per cápita y por puesto de trabajo  
(a precios constantes de 2000)**



Una vez enlazada, la serie de VAB “real” se expresa en precios del año 2000 multiplicándola por el nivel de precios en 2000 en base 95. El Gráfico 11 muestra la tasa de crecimiento anual de esta serie tras dividirla por la población y por el empleo. Llama la atención la persistente tendencia hacia la desaceleración del crecimiento de la productividad que se observa desde los primeros años sesenta hasta nuestros días.

## 6. Series regionales de empleo y VAB

En esta sección se construyen series enlazadas de empleo y VAB (a coste de los factores/precios básicos) medido a precios corrientes y constantes de 2000 para el período 1955-2007. El ejercicio se realiza en varias etapas. En primer lugar, se construye una serie “completa” basada en la CRE86 para el período 1980-96, rellenando los “huecos” que existen en la versión de esta serie que está disponible en la página web del INE a partir de otra información complementaria procedente generalmente del propio Instituto. Seguidamente, esta serie se combina con la serie del BBVA para obtener una *serie de referencia* para el período 1955-95 que es la que finalmente se enlaza con la serie terminal de CRE00-r para obtener las series enlazadas definitivas. La serie de referencia se construye extendiendo las series del BBVA hacia adelante con las tasas de crecimiento de las correspondientes series de la CRE86 a partir de un punto de corte que se determina por el procedimiento desarrollado en la sección 3.1. De esta forma, la serie de referencia conserva las tasas de crecimiento de la serie considerada más fiable (entre las del BBVA y la CRE86 durante su período de solapamiento) y los niveles originales de las series del BBVA para los primeros años de la muestra. Una vez enlazados el empleo y el VAB a precios

corrientes, la segunda serie se deflacta utilizando una serie de índices de precios que se obtiene por retropolación a partir de los deflatores de la CRE00-r y de la serie de referencia.

Exceptuando la primera y la última etapa, la mayor parte de los cálculos necesarios se realiza con series normalizadas, expresadas en términos relativos al promedio nacional por habitante. De esta forma, se neutralizan las posibles diferencias de nivel entre las distintas series y se utiliza únicamente la información que éstas contienen sobre la posición relativa de cada territorio. Las series de VAB a precios corrientes y de empleo de las distintas fuentes utilizadas se expresan en primer lugar en términos per cápita, dividiendo cada una de ellas por la serie enlazada "homogénea" de población de derecho (no corregida) descrita en la sección 4.1 y se transforman a continuación en índices, normalizando cada magnitud por su promedio nacional en el mismo año (excluyendo a la extra-región). Los enlaces del empleo y el VAB a precios corrientes se realizan en primera instancia con las series así normalizadas, recuperándose después los niveles por habitante y los valores totales de cada magnitud a partir del correspondiente promedio nacional (en las series agregadas enlazadas descritas en la sección 5) y de la población (de derecho) de cada región.

#### **6.1. Construcción de la serie completa de CRE86 para 1980-95**

Las series de empleo, población y VABcf a precios corrientes de la CRE86 que ofrece el INE en su página web comienzan en 1986. Para construir las series completas para 1980-95 que se confrontan con las del BBVA, he extendido estas series hacia atrás utilizando la serie homogénea 1980-89 en base 1986 publicada en su día por el INE.<sup>13</sup> En el caso del empleo, las dos fuentes coinciden en aquellos años comunes en las que la serie más antigua ya tiene carácter definitivo (1986 y 1987), por lo que basta con yuxtaponer ambas series, tomando los datos correspondientes a 1980-85 de la más antigua y el resto de la más reciente. Por otra parte, el VAB a coste de los factores no está disponible en la serie 1980-89, aunque sí lo está el VAB a precios de mercado. Por tanto, el VABcf de la serie más reciente se extiende hacia atrás desde 1986 hasta 1980 por retropolación, utilizando la tasa de crecimiento del VAB a precios (corrientes) de mercado en la serie más antigua.

Finalmente, en la CRE86 no se ofrecen estimaciones del VAB a precios constantes. Hay dos formas rápidas de obtener un índice de precios aproximado para esta serie. La primera es utilizar el deflactor del PIB regional que se puede recuperar a partir de las series de PIB a precios corrientes y constantes que ofrece la misma fuente. La segunda es trabajar con las series de VABcf a precios corrientes y constantes que se ofrecen en la BD Mores (de Bustos et al, 2008), donde se parte de los deflatores nacionales por ramas de actividad que proporciona el propio INE para construir índices de precios regionales. Puesto que el deflactor del PIB de CRE86 muestra un comportamiento un tanto errático, he optado por utilizar el deflactor del VABcf de la BDMores, pasado de base 2000 a base 1986, para deflactar el VABcf a precios corrientes de la CRE86.

---

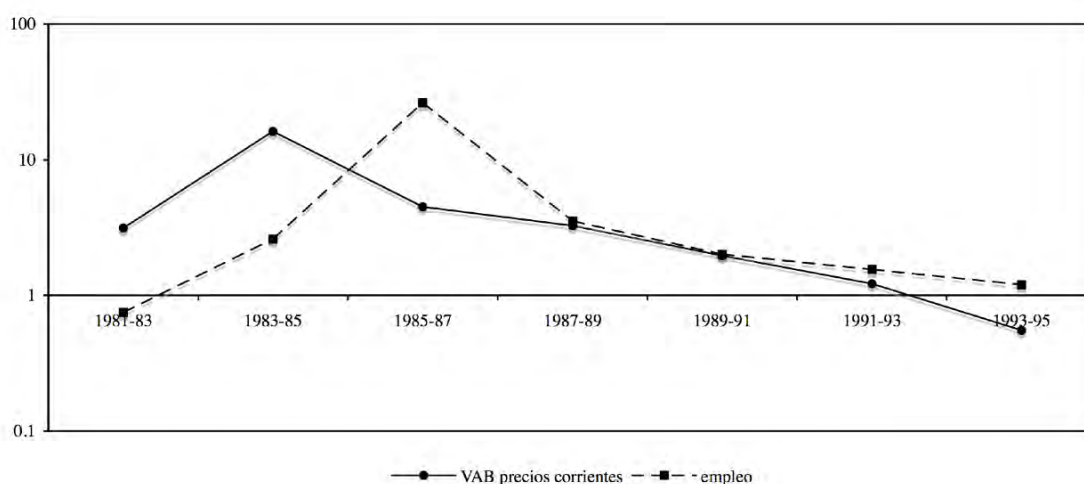
<sup>13</sup> Agradezco a Andrés de Bustos el haberme proporcionado estas series.

## 6.2. Construcción de las series de referencia de empleo y VABcf a precios corrientes para 1955-95

Las *series de referencia* para el período 1955-95 se construyen prolongando hacia adelante las series de empleo y VAB a precios corrientes de la Fundación BBVA utilizando las tasas de crecimiento de la correspondiente serie de CRE86. Para construir estas series, resulta necesario determinar la fecha óptima en la que abandonar las series del BBVA y adoptar como referencia la CRE86. En principio, esta fecha podría ser distinta para cada una de las series que estamos considerando. El método adoptado también permite la posibilidad de ir alternando las dos fuentes de acuerdo con los resultados que obtengamos en cada bienio. Sin embargo, he preferido excluir ex-ante estas dos posibilidades y elegir una única fecha de corte común para todas las series. El combinar series de VAB y empleo de fuentes diferentes parece arriesgado porque, aunque ninguna de las dos fuentes da demasiados detalles sobre sus procedimientos de estimación, uno tiende a pensar que las dos magnitudes no se estiman de forma independiente. Por otra parte, saltar de una serie a otra sería inconsistente con la intuición básica que guía el ejercicio: que hay problemas con los primeros años de la serie del INE pero que seguramente éstos se han ido corrigiendo con el paso del tiempo.

El punto de corte entre ambas fuentes se fija utilizando la metodología desarrollada en la sección 3.1. El cálculo de los ECMs de los estimadores obtenidos con las series del BBVA y de la CRE86 se realiza trabajando con las tasas de crecimiento anual (calculadas como diferencias logarítmicas) de las series estandarizadas. Puesto que los datos del BBVA sólo están disponibles en años alternos, he utilizado sólo los años impares de cada una de las series. Las primeras diferencias logarítmicas de la serie bienal se dividen por dos para obtener las tasas de crecimiento anualizadas con las que se realizan los cálculos.

**Gráfico 12: Varianza CRE/Varianza BBVA, series en tasas de crecimiento anualizadas (escala logarítmica)**

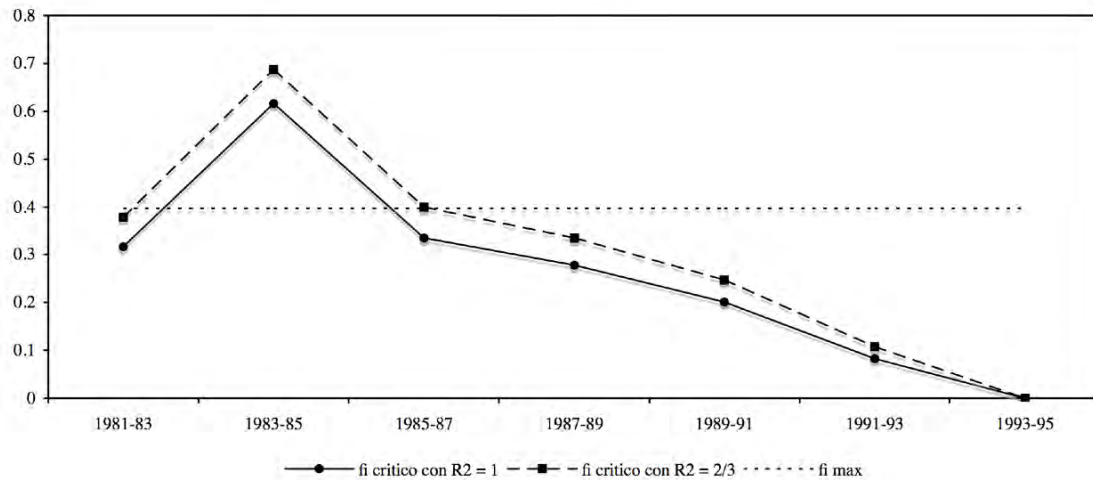


El Gráfico 12 muestra (en escala logarítmica) el ratio entre las varianzas de las tasas de crecimiento anualizadas del VAB y el empleo per cápita relativos en nuestras dos fuentes, con el INE en el numerador. El patrón temporal de las volatilidades relativas es similar al observado en el Gráfico 2 para el caso del PIB per cápita: con una única excepción puntual (el empleo en el

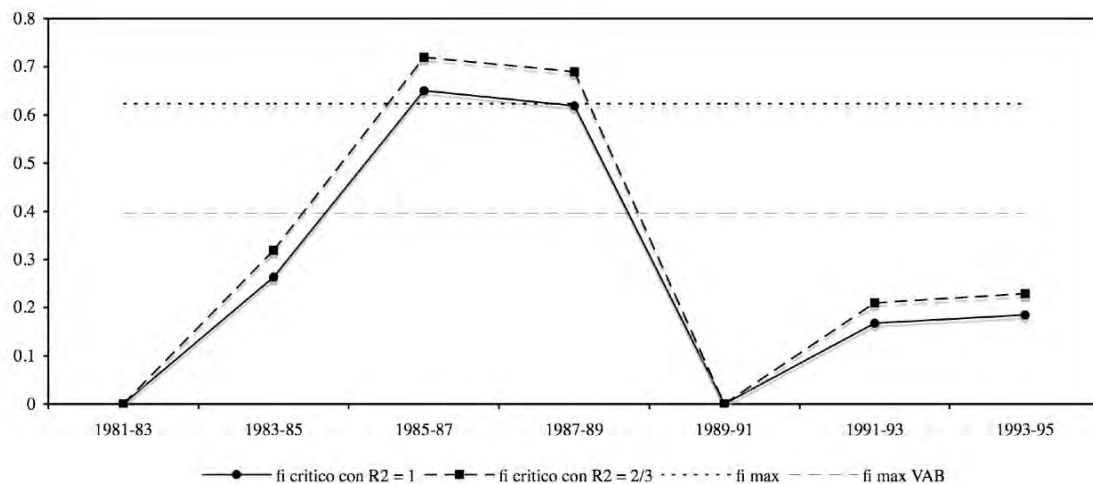
bienio 1981-83), las series de la CRE86 son mucho más volátiles que las del BBVA durante los primeros años del período y se acercan después gradualmente a estas últimas, llegando incluso a presentar varianzas inferiores en la parte final de la muestra.

**Gráfico 13: Valor crítico y cota superior de  $\phi$**

**a. VAB per cápita**



**b. Empleos per cápita**



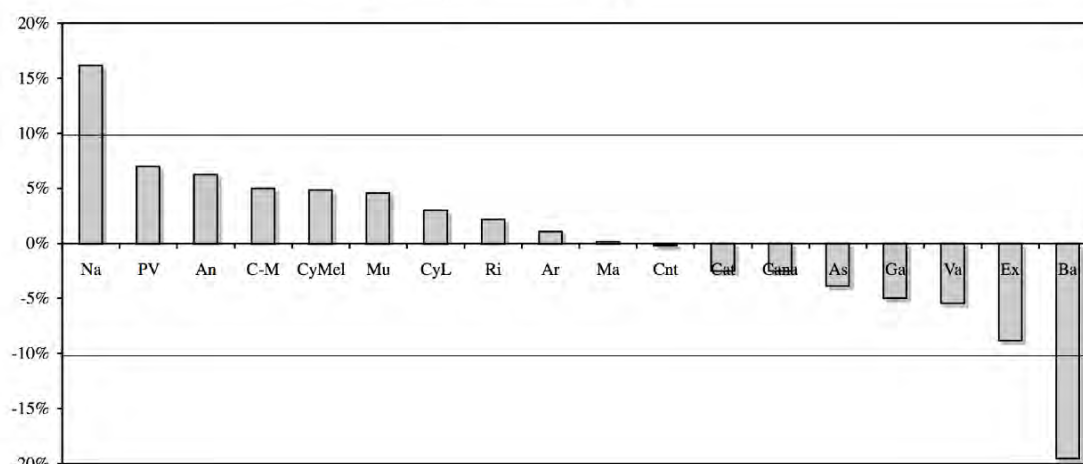
Como hemos visto en la sección 3, esto implica que las series del BBVA han de considerarse “mejores” que las de la CRE durante estos años siempre que no presenten un sesgo sistemático hacia la media demasiado elevado. El Gráfico 13 muestra los valores estimados del nivel crítico del parámetro que mide este sesgo,  $\tilde{\phi}$ , y de la cota superior para el mismo,  $\hat{\phi}_{\text{max}}$ . En el caso del VAB, los valores estimados de estos parámetros indican con claridad que deberían descartarse las tres primeras observaciones de la serie del INE (1981-87) y muy probablemente también la siguiente (1987-89). En el caso del empleo, los valores estimados de  $\tilde{\phi}$  se sitúan por encima de la cota superior en 1985-87 y 1987-89 pero no así en los dos bienios anteriores. Sin embargo, no parece probable que el sesgo sistemático de la serie del BBVA pueda ser mucho mayor en la estimación del empleo que en la del VAB. Si utilizamos para la serie de empleo la misma cota

de  $\phi$  que para la serie de VAB, la observación correspondiente a 1983-85 se sitúa cerca del umbral de rechazo. En conjunto, por tanto, parece razonable fijar la fecha de corte en 1989. En consecuencia, la serie de referencia se construye extendiendo las series del BBVA a partir de esta fecha hasta 1995, utilizando la tasa de crecimiento de la correspondiente serie de la CRE86, trabajando siempre con las variables estandarizadas por el promedio nacional por habitante.

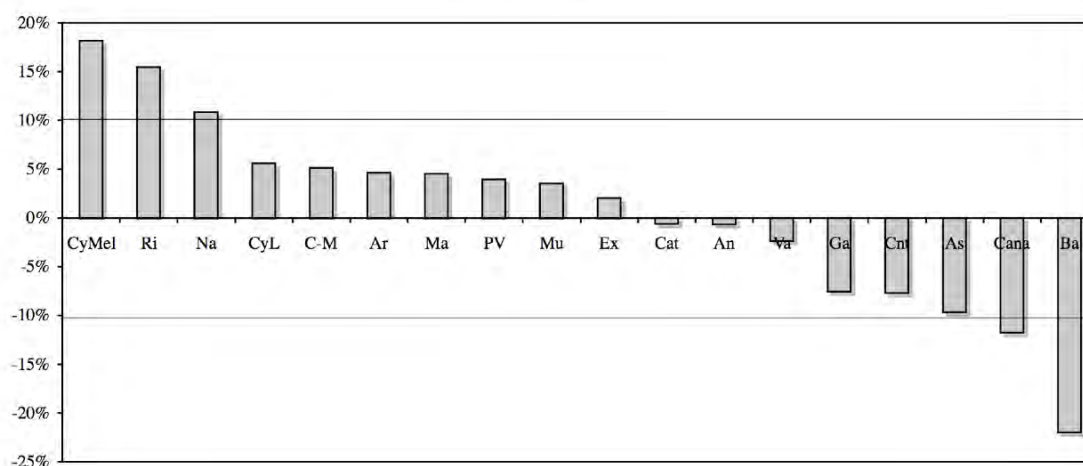
### 6.3. Discrepancias entre la serie de referencia y la serie terminal

El Gráfico 14 muestra la discrepancia existente en 1995 entre las series de referencia de VAB y empleo construidas en el apartado anterior y las correspondientes series terminales de la CRE00-r, con ambas medidas en logaritmos. En el caso del VAB, las diferencias son relativamente reducidas excepto en dos regiones: Baleares, que “pierde” casi veinte puntos de renta relativa al pasar de la serie de referencia a la serie terminal, y Navarra, que “gana” algo más de quince puntos porcentuales con el cambio. En el caso del empleo, el número de regiones en las que se observan discrepancias de más de 10 puntos porcentuales se eleva a cinco. Además de las dos ya citadas, se incorporan a la lista Ceuta y Melilla, la Rioja y Canarias.

**Gráfico 14: Diferencia porcentual en 1995 entre la serie terminal y la serie de referencia**  
a. VAB per capita relativo a precios corrientes



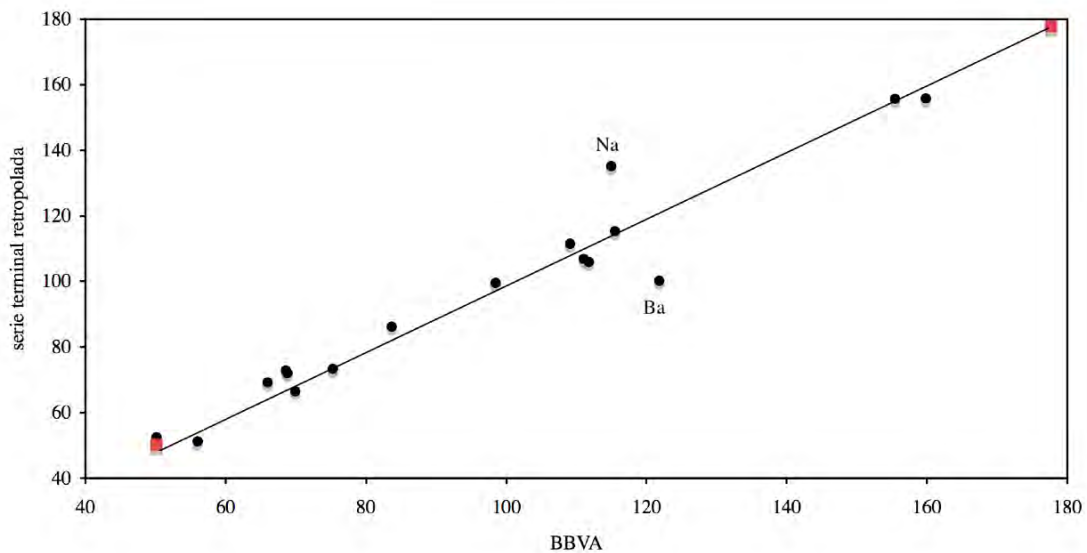
### b. Empleo per cápita relativo



- Nota:  $\ln(\text{CRE00-r}) - \ln(\text{serie de referencia})$

Suponiendo que las discrepancias porcentuales ilustradas en el Gráfico 14 se han mantenido constantes en el tiempo, podemos recuperar el valor correspondiente a 1955 de la serie terminal retropolada (STR) y compararlo con el valor correspondiente al mismo año de la serie de referencia (esto es, la estimación original de la Fundación BBVA para 1955). Los resultados de la comparación se resumen en los Gráficos 15 y 16 y los Cuadros 4 y 5.

**Gráfico 15: VAB a precios corrientes per cápita en 1955.  
Serie terminal retropolada vs. BBVA**

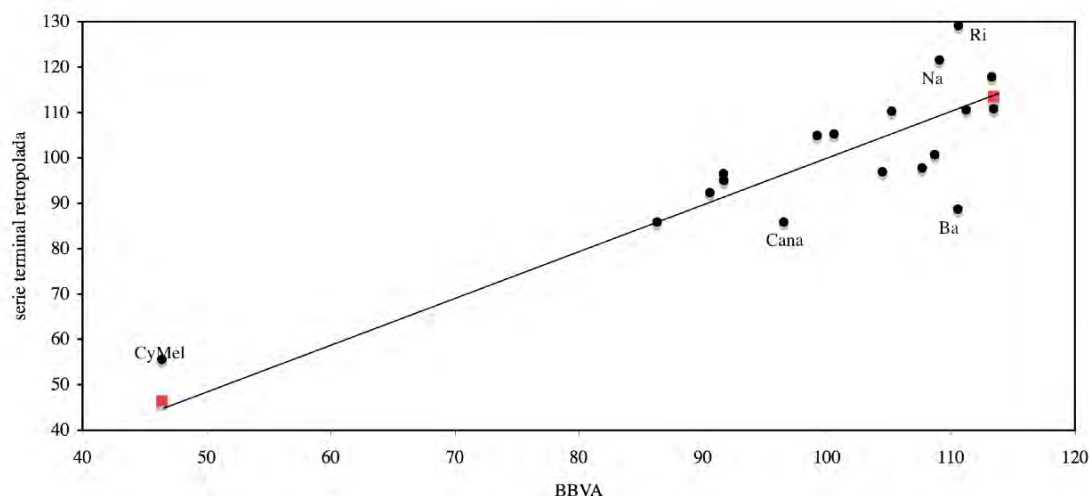


**Cuadro 4: VAB per cápita en 1955.  
Serie terminal retropolada vs BBVA**

	BBV	rank BBV	STR	rank STR	STR- BBVA	%	rankBBV- rank STR
<i>País Vasco</i>	177.7	1	190.6	1	12.9	7.2%	0
<i>Cataluña</i>	159.8	2	155.9	2	-3.9	-2.5%	0
<i>Madrid</i>	155.4	3	155.6	3	0.2	0.1%	0
<b><i>Baleares</i></b>	<b>121.8</b>	<b>4</b>	<b>100.3</b>	<b>9</b>	<b>-21.6</b>	<b>-17.7%</b>	<b>-5</b>
<i>Cantabria</i>	115.5	5	115.3	5	-0.2	-0.2%	0
<b><i>Navarra</i></b>	<b>115.0</b>	<b>6</b>	<b>135.2</b>	<b>4</b>	<b>20.2</b>	<b>17.6%</b>	<b>2</b>
<i>Valencia</i>	111.8	7	105.9	8	-5.9	-5.3%	-1
<i>Asturias</i>	111.1	8	106.9	7	-4.2	-3.7%	1
<i>Rioja</i>	109.1	9	111.5	6	2.4	2.2%	3
<i>Aragón</i>	98.4	10	99.5	10	1.1	1.1%	0
<i>Castilla y León</i>	83.7	11	86.2	11	2.6	3.1%	0
<i>Canarias</i>	75.3	12	73.4	12	-1.9	-2.5%	0
<i>Galicia</i>	69.9	13	66.5	16	-3.4	-4.8%	-3
<i>Murcia</i>	68.8	14	72.0	14	3.2	4.7%	0
<i>Andalucía</i>	68.6	15	73.0	13	4.4	6.4%	2
<i>Cast. la Mancha</i>	65.9	16	69.3	15	3.4	5.1%	1
<i>Extremadura</i>	56.0	17	51.3	18	-4.7	-8.4%	-1
<i>Ceuta y Melilla</i>	50.1	18	52.6	17	2.5	5.0%	1

- Nota: STR = serie terminal retropolada.

**Gráfico 16: Empleos per cápita en 1955.**  
Serie terminal retropolada vs. BBVA



**Cuadro 5: Empleos per cápita en 1955.**  
Serie terminal retropolada vs BBVA

	BBV	rank BBV	SEP	rank STR	SE- BBVA	%	rankBBV- rank SEP
<i>Valencia</i>	113.5	1	110.8	4	-2.6	-2.3%	-3
<i>País Vasco</i>	113.3	2	117.9	3	4.6	4.0%	-1
<i>Cataluña</i>	111.2	3	110.6	5	-0.6	-0.6%	-2
<b><i>Rioja</i></b>	<b>110.6</b>	<b>4</b>	<b>129.1</b>	<b>1</b>	<b>18.5</b>	<b>16.7%</b>	<b>3</b>
<b><i>Baleares</i></b>	<b>110.5</b>	<b>5</b>	<b>88.8</b>	<b>15</b>	<b>-21.8</b>	<b>-19.7%</b>	<b>-10</b>
<b><i>Navarra</i></b>	<b>109.1</b>	<b>6</b>	<b>121.6</b>	<b>2</b>	<b>12.5</b>	<b>11.4%</b>	<b>4</b>
<i>Cantabria</i>	108.7	7	100.7	9	-8.0	-7.4%	-2
<i>Asturias</i>	107.7	8	97.8	10	-9.9	-9.2%	-2
<i>Aragón</i>	105.2	9	110.2	6	5.0	4.7%	3
<i>Galicia</i>	104.5	10	96.9	11	-7.6	-7.2%	-1
<i>Madrid</i>	100.6	11	105.2	7	4.6	4.6%	4
<i>Castilla y León</i>	99.2	12	104.9	8	5.7	5.8%	4
<b><i>Canarias</i></b>	<b>96.5</b>	<b>13</b>	<b>85.9</b>	<b>16</b>	<b>-10.7</b>	<b>-11.0%</b>	<b>-3</b>
<i>Murcia</i>	91.7	14	95.0	13	3.3	3.6%	1
<i>Cast. la Mancha</i>	91.6	15	96.5	12	4.8	5.3%	3
<i>Extremadura</i>	90.5	16	92.4	14	1.9	2.1%	2
<i>Andalucía</i>	86.3	17	85.8	17	-0.5	-0.6%	0
<b><i>Ceuta y Melilla</i></b>	<b>46.4</b>	<b>18</b>	<b>55.6</b>	<b>18</b>	<b>9.2</b>	<b>19.9%</b>	<b>0</b>

- Nota: STR = serie terminal retropolada.

Un resultado alentador del ejercicio es que los valores iniciales de las dos series que estamos comparando no son demasiado distintos en términos estadísticos. La correlación entre la serie terminal retropolada y la serie de referencia en 1955 es de 0.976 en el caso del VAB y de 0.828 en el del empleo. Para la mayoría de las regiones, por tanto, el enlace entre las dos series (por cualquier procedimiento) no exige una revisión radical de la historia. Como ya hemos visto, sin embargo, hay algunas excepciones. El caso más llamativo es el de Baleares. Tomando como referencia la estimación del BBVA, la retropolación de la serie terminal reduce la renta per cápita relativa del archipiélago en 21.6 puntos y su empleo por habitante relativo en 21.8



puntos. En términos de la primera variable, la región pasaría de la cuarta a la novena posición y en términos de la segunda, de la quinta a la decimoquinta. También son preocupantes las discrepancias sobre Navarra (20.2 puntos en términos de renta y 12.5 en términos de empleo), la Rioja (18.5 puntos en términos de empleo) y Canarias y Ceuta y Melilla (10.7 y 9.2 puntos de empleo).

#### **6.4. Comparación con fuentes externas y selección del valor de $\rho$**

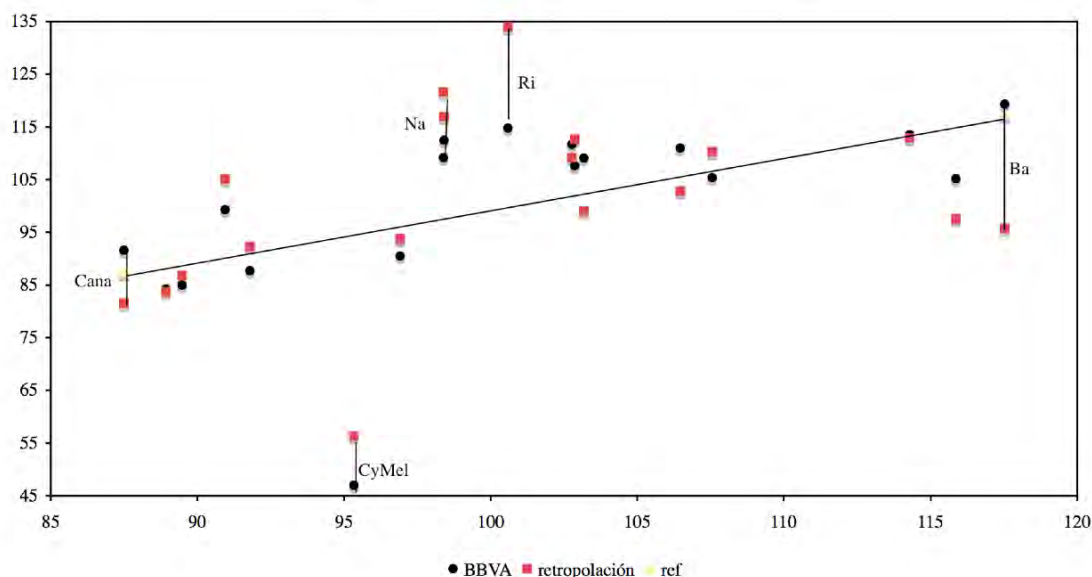
El problema con el que nos enfrentamos a la hora de construir el enlace entre las dos series con las que estamos trabajando es el de cómo repartir la diferencia entre ambas que “aflora” en 1995 entre el nivel inicial y las tasas de crecimiento subsiguientes de la serie de referencia. Para intentar basar esta decisión en algo más firme que la mera conjetura, he intentado buscar fuentes “externas” de información que nos puedan dar alguna idea de la fiabilidad relativa de las dos estimaciones alternativas de las que disponemos para la renta y el empleo iniciales: la estimación de la Fundación BBVA para algún año cercano al comienzo del período muestral y el valor correspondiente al mismo año de la serie terminal retropolada.

##### ***Empleo por habitante***

En el caso del empleo, el Censo de 1960 nos proporciona una estimación externa en principio fiable de la tasa de ocupación relativa de las regiones españolas al comienzo del período muestral. Un problema con esta fuente es que en la misma no se identifica a los parados, que parecen incluirse en una categoría residual dentro de la población activa. Por lo tanto, el número de parados en 1960 (calculado como el promedio de los registrados a finales de los meses de junio y de diciembre) se toma del *Anuario Estadístico de España* de 1961. Substrayendo esta cifra de la población activa censal, se obtiene la población ocupada, que se divide por la población de derecho y se normaliza por el promedio nacional para obtener el dato más cercano posible a las series de empleo (puestos de trabajo) con las que estamos trabajando. El dato censal así obtenido se compara con una estimación del empleo per cápita relativo de cada región en 1960 de acuerdo con la serie del BBVA y con la serie terminal retropolada, que se obtiene como el promedio de los valores correspondientes a 1959 y 1961 de cada una de ellas.

El Gráfico 17 muestra los resultados. La serie retropolada y la serie del BBVA se miden en escala vertical, mientras que el dato censal se mide en el eje horizontal. Puesto que la recta que aparece en el gráfico corresponde a la diagonal o línea de  $45^\circ$ , la distancia vertical entre cada punto y esta recta mide la diferencia entre cada una de las series de interés y el dato censal correspondiente. En términos generales, la serie del BBVA se parece más al dato censal que la serie retropolada. Los puntos negros que representan a la primera están más cerca de la recta de  $45^\circ$  que los cuadrados rojos que corresponden a la serie retropolada en la mayoría de las regiones. De la misma forma, la correlación con el dato censal (excluyendo a Ceuta y Melilla) es mucho más alta en el caso del BBVA (0.764) que en el de la serie enlazada (0.341).

**Gráfico 17: Comparación con el censo de 1960.  
Serie de empleo del BBVA y serie terminal retropolada**



En principio, por tanto, parece razonable asignar un peso superior a la estimación del BBVA que a la serie retropolada, esto es, fijar un valor reducido para  $\rho$  a la hora de realizar el enlace por el método mixto. En consecuencia, he asignado un valor de 0.10 a este parámetro.

#### *Ceuta y Melilla*

Ceuta y Melilla constituyen un caso muy atípico. Como se observa en el Gráfico 17, tanto la retropolación de la serie terminal de empleo como la estimación del BBVA toman un valor muy inferior (entre 30 y 40 puntos) al dato censal de ocupación, lo que aconseja examinar en mayor detalle la información disponible sobre la evolución del empleo en este territorio.

**Cuadro 6: Ceuta y Melilla, empleos por habitante relativos, diversas fuentes  
(España = 100)**

	1960	1970	1980	1990
<i>Censo</i>	95.3	79.8		78.2
<i>BBVA</i>	47.0	58.3	74.1	84.4
<i>CRE86</i>			75.9	85.7

- *Nota:* los datos del BBVA para años censales se obtienen promediando los valores correspondientes al año anterior y posterior a cada censo. Los censos de 1960 y 1970 contienen información sobre la población activa pero no sobre el número de parados. Este último dato se toma del *Anuario Estadístico de España*. Los datos censales y los de la CRE86 corresponden al número de personas ocupadas por habitante de derecho, mientras que los del BBVA se basan en el número de puestos de trabajo.

El Cuadro 6 muestra la información que he podido reunir sobre la evolución del empleo relativo de las dos ciudades autónomas en años seleccionados (medido por el número de ocupados o por el número de puestos de trabajo por habitante de derecho en relación al promedio nacional). Los datos censales sugieren un fuerte descenso del empleo relativo entre 1960 y 1970, seguido de una estabilización a un nivel inferior en unos veinte puntos al promedio nacional. La serie del BBVA, por contra, indica un incremento fuerte y sostenido de esta

variable durante todo el período muestral a partir de un nivel inicial muy inferior al promedio nacional. Puesto que la estimación del BBVA coincide aproximadamente con la de la CRE86 a partir de 1980 y no está muy lejos del dato censal de 1990, parece razonable conservarla como serie de referencia del primer año citado en adelante. En la primera parte del período muestral, sin embargo, la fuerte discrepancia con el dato censal aconseja algún tipo de corrección al alza.

En consecuencia, he optado por revisar la serie de referencia de empleo para Ceuta y Melilla antes de proceder a su enlace con la serie terminal de la CRE00-r. La nueva serie de referencia se construye como sigue. De 1981 en adelante, se mantiene la serie original del BBVA, prolongada con las tasas de crecimiento de la CRE86 a partir de 1989. Para 1960 y 1970 la nueva serie se obtiene como una media ponderada de las estimaciones del BBVA (con peso 2/3) y del correspondiente dato censal (con peso 1/3). El dato de 1960 se extiende hacia atrás hasta 1955 por retropolación, utilizando las tasas de crecimiento de la serie del BBVA. Entre 1960 y 1970 la serie se completa respetando en lo posible el perfil temporal de la serie original del BBVA, pero reescalando sus incrementos de forma que la serie pase por los puntos deseados en los años citados. Esto es, para cada  $t$  entre 1960 y 1970 construimos

$$(32) \Delta_t = \frac{EMBBVA_t - EMPBBVA_{1960}}{EMPBBVA_{1970} - EMPBBVA_{1960}}$$

de forma que  $\Delta_t$  mide la fracción del incremento total del empleo durante la década de interés que se ha producido ya en el año  $t$ . El valor en  $t$  de la nueva serie de referencia de empleo se estima como

$$(33) EMREF_t = EMREF_{1960} + \Delta_t (EMREF_{1970} - EMREF_{1960})$$

El mismo procedimiento se utiliza también para completar la serie ente 1970 y 1981, año éste en el que se vuelve a la serie original del BBVA. La serie así construida reemplaza a la serie original del BBVA como serie de referencia entre 1955 y 1981 y se enlaza con la serie terminal utilizando un valor de 0.10 para  $\rho$ .<sup>14</sup>

### *VAB por habitante*

En el caso del VAB resulta más complicado encontrar una estimación verdaderamente externa del valor de esta magnitud cerca del comienzo del período muestral. Alvarez Llano (1986) y Martín Rodríguez (1992 y 1993) ofrecen estimaciones del nivel de renta de las regiones españolas en 1950 o 1960 como parte de ambiciosos intentos por aproximar la evolución económica de estos territorios durante un período de varios siglos. Para los años que nos interesan, sin embargo, ambos autores se apoyan en datos de la Fundación BBVA o en trabajos de Alcaide relacionados con ellos. La única estimación aparentemente independiente de esta última fuente que he podido encontrar es la que construye Plaza Prieto (1953) para el año 1949 utilizando procedimientos indirectos para provincializar por ramas una estimación del

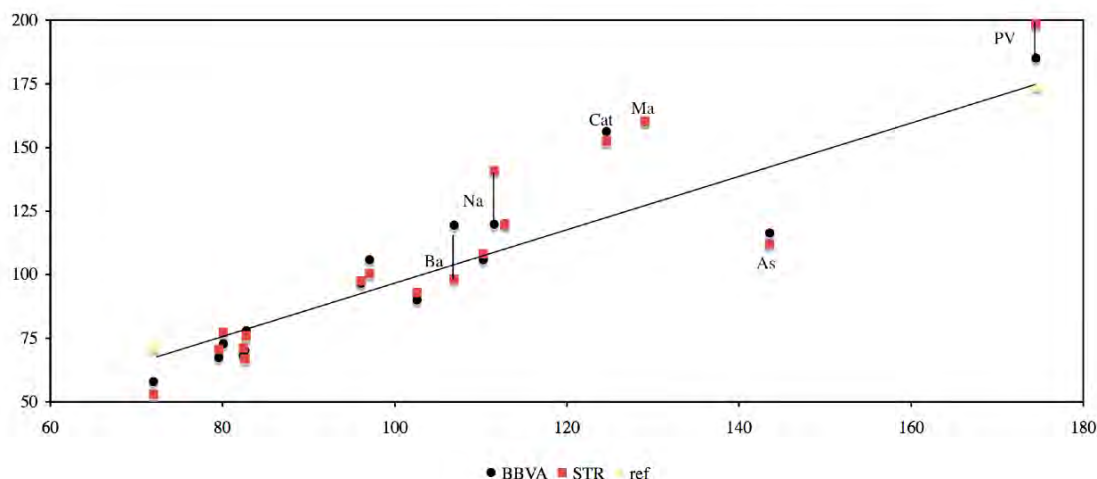
---

<sup>14</sup> No hace falta insistir en que la serie de empleo para Ceuta y Melilla que aquí se ofrece es de dudosa fiabilidad en sus primeros años. En aplicaciones, podría ser prudente excluir a este territorio de la muestra, especialmente si constituye una observación atípica en materia de empleo.

producto agregado español. Dadas las enormes limitaciones de la información disponible, existen serias dudas sobre la fiabilidad de la estimación de Plaza. Sin embargo, cabe esperar que estas limitaciones sean algo menos importantes a la hora de aproximar la situación relativa de cada uno de los territorios y, en cualquier caso, parece tratarse de la única fuente externa existente.

Las estimaciones del BBVA y la retropolación de la serie terminal se comparan con los datos de Plaza Prieto en el Gráfico 18, que se construye de la misma forma que el Gráfico 17, con la serie de Plaza medida en el eje horizontal y las otras dos en el eje vertical.<sup>15</sup> En primer lugar, hay que resaltar que la correlación entre las estimaciones de este autor y las otras dos series es muy elevada y adopta un valor similar en ambos casos: 0.914 con la serie del BBVA y 0.910 con la serie terminal retropolada. En algunos territorios, estas dos últimas series se alejan significativamente de la estimación de Plaza pero, en términos generales, coinciden entre sí cuando esto sucede. Los casos más problemáticos son los del País Vasco, Navarra, Extremadura y Baleares (véase el Cuadro 4). En los tres primeros territorios, la estimación del BBVA está más cercana a la de Plaza, mientras que en el último sucede lo contrario.

**Gráfico 18: Comparación con la estimación de Plaza Prieto.  
Serie de VAB del BBVA y serie terminal retropolada, 1950**



La información disponible, por tanto, sugiere que la serie del BBVA sobrevalora significativamente la renta de Baleares a mediados del siglo pasado,<sup>16</sup> pero no necesariamente infravalora la de las otras tres regiones citadas. A la vista de estos resultados, el valor de  $\rho$  se ha fijado como sigue. Como criterio general, he elegido un valor de 0.5, lo que supone dar el mismo peso a la serie del BBVA y a la retropolación de la serie terminal a la hora de estimar el

<sup>15</sup> La serie de VAB del BBVA se extiende hacia atrás desde 1955 hasta 1950 utilizando la tasa de crecimiento de la misma variable durante este período de acuerdo con Alcalde (2007), donde se ofrecen estimaciones de renta regional a intervalos quinquenales para todo el siglo XX. El valor de la serie terminal retropolada en el mismo año se obtiene añadiendo al dato estimado del BBVA en 1950 la diferencia logarítmica entre la serie del BBVA y la serie terminal en 1995. El dato de Alcalde (2007) para 1950 no se utiliza directamente porque existen pequeñas discrepancias entre esta fuente y las series del BBVA en 1955.

<sup>16</sup> Maluquer (2006) también considera muy poco plausibles las estimaciones de renta la FBBVA para Baleares.

nivel inicial de renta. Las excepciones son el País Vasco, Extremadura y Navarra, para las que he fijado  $\rho = 1/3$ , otorgando así un peso mayor a la estimación del BBVA que a la retropolación de la serie terminal, y Baleares, para la que utilizaré  $\rho = 2/3$ , invirtiendo las ponderaciones de ambas fuentes.

#### 6.5. Enlace de las series de empleo y VAB a precios corrientes

El enlace entre las series terminales (CRE00-r) y de referencia (BBVA+CRE86) de empleo y VAB a precios corrientes se realiza por el procedimiento mixto, utilizando los valores de  $\rho$  indicados en la sección anterior y resumidos en el Cuadro 7. El resultado obtenido por este procedimiento se renormaliza de forma que el promedio nacional (ponderado por población) calculado a partir de las nuevas series regionales enlazadas sea exactamente igual a 100 en cada período.

Una vez obtenidas las series definitivas de VAB y empleo por habitante en términos relativos, los niveles de VAB y empleo por habitante se recuperan aplicando los índices enlazados al correspondiente promedio nacional tomado de la sección 5. La cifra resultante se multiplica a su vez por la población de derecho regional para recuperar el empleo total de cada territorio (en puestos de trabajo) y su VAB a precios básicos (medido en euros corrientes).

**Cuadro 7: Valores de  $\rho$  utilizados para enlazar las distintas series**

	VAB	empleo
<i>caso general</i>	0.50	0.10
excepciones:		
<i>País Vasco</i>	1/3	
<i>Navarra</i>	1/3	
<i>Extremadura</i>	1/3	
<i>Baleares</i>	2/3	

#### 6.6. Deflatores y series de VAB a precios constantes

Los deflatores del VAB implícitos en las series del BBVA se obtienen dividiendo la serie de VAB a precios corrientes por la correspondiente serie a precios constantes (de 1986). En FBBV (1999, pp. 36-37) se dan algunas indicaciones sobre la construcción de las series de precios. En el caso de la industria y los servicios, el VAB a precios corrientes de cada rama productiva se deflata en todas las regiones utilizando el índice de precios nacional que proporciona la CNE (para 21 ramas productivas). En estos sectores, por tanto, las diferencias de niveles de precios entre regiones de acuerdo con las series del BBVA recogen únicamente diferencias en su estructura productiva. En el caso de la agricultura, sin embargo, se parte de los datos que ofrece el Ministerio de Agricultura sobre los precios de diversos productos y la producción de los mismos en cada región para construir un índice específico para cada territorio. También se construye un índice de precios específico para el sector de la construcción, pero no se ofrece detalle alguno sobre la metodología utilizada, remitiendo al lector a un trabajo que no he conseguido encontrar (Alcaide y Pascualena, 1985).

Como hemos visto en un apartado anterior, el índice de precios del VAB que he utilizado en la serie completa de CRE86 está tomado de la BDMores00 y se construye también a partir de los deflatores sectoriales nacionales. Por último, en la CRE00 ya no se ofrecen series de VAB a precios constantes sino índices enlazados de volumen con base 2000 (esto es, con un valor normalizado a 100 en el año 2000 para todas las regiones). Estas series seguramente recogen mejor que sus antecesoras a precios constantes las variaciones del producto real de cada región a lo largo del tiempo pero no contienen información alguna sobre sus niveles. Dado que esta última es necesaria en muchas ocasiones, he reconstruido una serie de VAB "a precios constantes de 2000" (lo que estrictamente hablando no es cierto) multiplicando los índices de volumen regionales por el VAB a precios corrientes de cada región en el año 2000. Para evitar problemas de no-aditividad, la serie nacional a precios constantes se obtiene sumando las magnitudes regionales (y no a partir del índice de volumen agregado). Dividiendo la serie de VAB a precios corrientes por la serie de VAB a precios constantes así construida, se obtiene un índice aproximado de precios del VAB con base 2000.<sup>17</sup>

Estas tres series de deflatores se enlazan por retropolación, extendiendo desde 1995 hacia atrás la serie de precios del VABpb de la CRE00 utilizando las tasas de crecimiento de las series de precios del VAB de CRE86 (de hecho, de la BDMores) y del BBVA, manteniendo el año de corte entre ambas fuentes fijado en una sección anterior (1989).<sup>18</sup> En esta ocasión el enlace no se realiza con las series normalizadas por el promedio nacional sino con índices "brutos" de precios para cada región.

Una vez construida una serie preliminar de precios por este procedimiento, esta serie se utiliza para deflactar la serie enlazada de VABpb nominal, obteniendo una serie preliminar de VAB real valorada (aproximadamente) a precios constantes de 2000. Esta serie preliminar se ajusta proporcionalmente de forma que su suma coincida con la serie nacional de VABpb a precios constantes construida en la sección 5. Dividiendo la serie enlazada de VAB a precios corrientes por esta serie final de VAB a precios constantes, se recupera la serie final de precios en base 2000.

### *Productividad a precios medios del período muestral*

En ocasiones nos interesa comparar el *nivel* de productividad de distintas regiones además de su tasa de crecimiento. El ejercicio, sin embargo, puede ser problemático cuando se realiza con datos a precios constantes porque los niveles relativos de productividad regional pueden ser muy sensibles al año que se tome como base en la serie de precios. El Gráfico 19 ilustra el

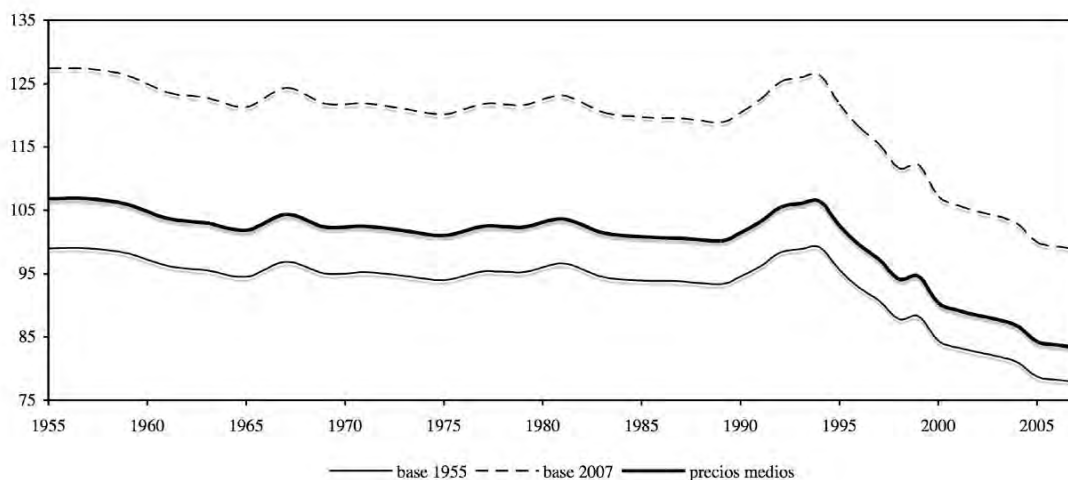
---

<sup>17</sup> En la mayor parte de los casos, los deflatores de la CRE00 son específicos para cada región incluso a nivel de ramas, pero toman como base el nivel de precios de cada territorio en 2000, con lo que no tienen en cuenta las posibles diferencias entre regiones en niveles de precios en ese año.

<sup>18</sup> El criterio estadístico utilizado más arriba para valorar el contenido informativo de las distintas series no parece aplicable en el caso de los índices de precios. Dado que en las fuentes que manejamos estos índices se construyen fundamentalmente a partir de los deflatores nacionales por ramas, hemos de esperar que su volatilidad subestime la de los precios "reales" dado que sólo se recogen las variaciones medias nacionales dentro de cada rama. En este contexto, una menor volatilidad no es necesariamente una buena señal. Dada la dificultad de encontrar un criterio razonable para intentar valorar estas series, he optado por respetar el punto de corte fijado arriba (y por no incluir a las series de precios en el análisis realizado para fijarlo).

problema con datos de Baleares. Puesto que la economía de esta región es muy intensiva en servicios y los precios de los servicios han crecido bastante más que los de los bienes industriales o agrícolas, Baleares aparenta ser mucho más productiva si su VAB por puesto de trabajo se valora a precios constantes del final del período muestral que cuando se hace con los precios del comienzo del mismo.

**Gráfico 19: Producto por puesto de trabajo, Baleares  
a precios constantes de distintos años y a precios medios del período muestral**



Para intentar mitigar este problema, he construido series de VAB valoradas a precios medios del período muestral. Esto es, en vez de dividir la serie de VAB a precios corrientes por un índice de precios con base en un año determinado, la serie se deflacta utilizando la media geométrica de los distintos índices de precios que se obtendrían tomando como base cada uno de los años impares incluidos en el período muestral.<sup>19</sup>

La construcción del deflactor apropiado se realiza como sigue. Sea  $P_{it}^{BT}$  el deflactor del VAB de la región  $i$  correspondiente al año  $t$  en base  $T$  ( $= 2000$  en nuestro caso). Para cambiar la referencia de la serie de precios, obteniendo así un índice "en base  $S$ " tenemos que dividir el índice original de precios en base  $T$  por el valor correspondiente (también en base  $T$ ) al nuevo año de referencia,  $S$ :

$$(34) P_{it}^{BS} = \frac{P_{it}^{BT}}{P_{is}^{BT}}$$

o, utilizando minúsculas para indicar que estamos trabajando con logaritmos,

$$(35) p_{it}^{BS} = p_{it}^{BT} - p_{is}^{BT}$$

El índice de precios que queremos calcular es la media geométrica de los que se obtendrían tomando como base cada uno de los años (impares) del período muestral, esto es,

<sup>19</sup> Se toman sólo los años impares porque la serie enlazada tiene periodicidad bienal hasta 1989. Utilizar todas las observaciones supondría ponderar más la parte final del período que la inicial.

$$(36) \bar{P}_{it} = \left( \prod_{n=1}^N P_{it}^{Bn} \right)^{1/N}$$

donde  $N$  es el número de observaciones por región (correspondientes a años impares). Tomando logaritmos de esta expresión y utilizando (35), tenemos

$$(37) \bar{P}_{it} = \frac{1}{N} \sum_n P_{it}^{Bn} = \frac{1}{N} \sum_n (P_{it}^{BT} - P_{in}^{BT}) = P_{it}^{BT} - \frac{1}{N} \sum_n P_{in}^{BT}$$

Obsérvese que esta ecuación nos permite calcular  $\bar{P}_{it}$  directamente a partir de los índices de precios en base 2000 ( $=T$ ), sin necesidad de construir explícitamente todas las series de precios con base en cada uno de los años impares del período muestral. Obsérvese también que el resultado no depende del año que se tome como base para la serie de precios. Si partimos de una serie de precios con referencia  $S$ , tendremos

$$(38) \bar{P}_{it}^* = P_{it}^{BS} - \frac{1}{N} \sum_n P_{in}^{BS} = (P_{it}^{BT} - P_{is}^{BT}) - \frac{1}{N} \sum_n (P_{in}^{BT} - P_{is}^{BT}) = P_{it}^{BT} - \frac{1}{N} \sum_n P_{in}^{BT} = \bar{P}_{it}$$

donde he utilizado (35) para pasar los índices de precios de base  $S$  a base  $T$ .

Como en el caso anterior, las series regionales de VAB a precios medios se agregan para recuperar el total nacional y el deflactor nacional se obtiene dividiendo la serie nacional a precios corrientes por la serie a precios medios.

## 7. Archivos de datos y gráficos comparativos

El trabajo va acompañado de dos archivos de Excel. El primero (*REGDAT\_v21\_5507.xls*) contiene las nuevas series enlazadas y el segundo (*datos\_originales.xls*) todos los datos de otras fuentes que se han utilizado para construir tales series. Los Cuadros 8 y 9 resumen el contenido de ambos archivos.

**Cuadro 8: Contenido del archivo de datos *REGDAT\_v21\_5507.xls***

---

<b>Series nacionales (excluyendo la extra-regio)</b>
- VAB a coste de los factores/precios básicos, a precios corrientes y de 2000
- Empleo (puestos de trabajo)
- Población residente a 1 de julio
<b>Series regionales</b>
- VAB a coste de los factores/precios básicos, a precios corrientes, constantes de 2000 y precios medios del período 1955-2007
- Empleo (puestos de trabajo)
- Población residente a 1 de julio
- Población de hecho a 1 de julio
- Población residente ajustada a 1 de julio

---



## Cuadro 9: Contenido del archivo de datos *datos\_originales.xls*

---

### **hoja 1. Agregados nacionales, 1954-96**

- Producto Interior Bruto a precios de mercado a precios corrientes y constantes de 1986
- Valor Añadido Bruto a precios de mercado a precios corrientes y constantes de 1986
- Valor Añadido Bruto a coste de los factores a precios corrientes y constantes de 1986
- Producción imputada de servicios bancarios
- VAB a coste de los factores de la "extra-regio"
- Empleo total (ocupados)
- Población de derecho o residente a 1 de julio
- Población de hecho a 1 de julio

*Fuentes:* INE: CNE y CRE base 1986, estimaciones intercensales de población y estimaciones de la población actual; Uriel, Moltó y Cuarella (2000), Maluquer y Llonch (2005), Maluquer (2008)

### **hoja 2: Series regionales de población del INE**

- Población de derecho o residente a 1 de julio
- Población de hecho, censos de 1950, 60, 70, 81 y 91
- Residentes ausentes como % de la población de derecho, censos de 1950, 60, 70, 81 y 91

*Fuentes:* INE, Cifras de población: [http://www.ine.es/inebmenu/mnu\\_cifraspob.htm](http://www.ine.es/inebmenu/mnu_cifraspob.htm)

### **hoja 3: Series regionales de la Fundación BBVA**

- Valor Añadido Bruto a coste de los factores a precios corrientes y constantes de 1986
- Empleo total (puestos de trabajo)
- Población residente a 1 de julio
- Producto Interior Bruto

*Fuentes:* Fundación BBV (1999) y Fundación BBVA (2000).

### **hoja 4: CRE00-r**

- Valor Añadido Bruto a precios básicos a precios corrientes y constantes de 2000
- Empleo total (puestos de trabajo)

*Fuentes:* INE, CRE base 2000 (para 2000 y años posteriores) y de la Fuente (2009) para 1995-1999.

### **hoja 5: CRE86**

- Producto Interior Bruto a precios corrientes y constantes de 1986
- Valor Añadido Bruto a coste de los factores a precios corrientes
- Empleo total (ocupados)
- Población de derecho
- Valor Añadido Bruto a precios de mercado a precios corrientes
- Empleo total (ocupados)
- Población de derecho

*Fuentes:* INE, CRE base 1986

### **hoja 6: BDMORES00**

- VAB a coste de los factores/precios básicos a precios corrientes y constantes de 2000
- Empleo total (ocupados/puestos de trabajo)
- Población de derecho

*Fuente:* de Bustos et al (2008)

### **hoja 7: CRE95**

- Valor Añadido Bruto a precios básicos, a precios corrientes y constantes de 1995
- Empleo total (puestos de trabajo)

*Fuente:* INE, CRE base 1995

---

### Cuadro 9: Contenido del archivo de datos *datos\_originales.xls* – continuación

---

#### hoja 8: Enlace INE de la CRE 1995-00

- VAB a precios básicos a precios corrientes y en índices de volumen con referencia 2000
- Empleo total (puestos de trabajo)

Fuente: INE, CRE Base 2000. Serie homogénea 1995-2008

---

En el Anexo 2 las series enlazadas se comparan con las series correspondientes del BBVA, de la CRE86 y de la BDMores. Para cada región se construyen gráficos comparativos (Gráficos A2.1-A2.18) para tres variables: el VABpb/cf por habitante de derecho a precios corrientes, el empleo por habitante y el producto por puesto de trabajo medido a precios constantes de 2000. Para construir la última variable, todos los índices de precios se han puesto "en base 2000". Para ello, la serie de precios del BBVA en base 1986 se extiende hasta 2000 utilizando las tasas de crecimiento de la serie de precios de CRE00-r y luego se pone en base 2000. La evolución de la serie de productividad se muestra también valorada a precios de 1955, 2007 y precios medios del período en el Gráfico A2.19.

#### 8. Conclusión

En este trabajo se construyen series "homogéneas" largas de algunos agregados económicos regionales mediante el enlace de los datos de la Fundación BBVA con los de diversas versiones de la Contabilidad Regional. El "punto de corte" en el que la primera de estas fuentes se abandona como referencia para la construcción de la serie enlazada se determina utilizando un procedimiento que permite estimar cuál de las dos series disponibles generará un estimador con menor error cuadrático medio cuando se utiliza como variable dependiente en una regresión sobre una variable independiente arbitraria.

En la mayor parte de los casos, el enlace no es problemático en el sentido de que las discrepancias existentes entre las distintas series en el momento de su enlace son generalmente reducidas, al menos cuando se trabaja con series normalizadas por el promedio nacional por habitante. En los casos en los que esto no es cierto, he intentado apoyarme en evidencia externa a la hora de distribuir la diferencia aflorada en el momento del enlace entre los niveles iniciales de las series de la Fundación BBVA y sus tasas de crecimiento durante el período muestral. Aún así, ha de reconocerse que el procedimiento utilizado deja un cierto margen para decisiones subjetivas que son seguramente discutibles. Como contrapartida, tiene la ventaja de hacer explícita la necesidad de realizar algún tipo de hipótesis sobre la senda temporal del error aflorado, afrontando así de una forma transparente una dificultad que los métodos más habituales de enlace tienden a ocultar pero no resuelven.

## Referencias

- Alcaide, J. (2007). *Evolución de la población española en el siglo XX por provincias y comunidades autónomas*. Fundación BBVA, Bilbao.
- Alcaide, C. y R. Pascualena (1985). "Deflatores provinciales y regionales de las macromagnitudes económicas." Caja de Ahorros de Alicante y Murcia, Alicante.
- Alvarez Llano, R. (1986). "Evolución de la estructura regional de España en la historia. Una aproximación." *Situación 1*, pp. 5-61.
- Cañada, A. (1995). "Algunos aspectos metodológicos de la estimación del empleo en la Contabilidad Nacional de España." *Estadística Española* 37(138), pp. 45-73.
- de Bustos, A., A. Cutanda, A. Díaz, F. J. Escribá, M. J. Murgui y M. J. Sanz (2008). "La BD MORES en base 2000: nuevas estimaciones y variables." Documento de Trabajo D2008-02, Dirección General de Presupuestos, Ministerio de Hacienda. Madrid.  
[http://www.igae.meh.es/SGPG/Cln\\_Principal/Presupuestos/Documentacion/Basesdatos\\_estudiosregionales.htm](http://www.igae.meh.es/SGPG/Cln_Principal/Presupuestos/Documentacion/Basesdatos_estudiosregionales.htm)
- de la Fuente, A. (2008). "Series enlazadas de algunos agregados económicos regionales, 1995-2007. Versión 1.1." Documento de Trabajo D2008-03. Dirección General de Presupuestos, Ministerio de Economía y Hacienda, Madrid.  
<http://ideas.repec.org/p/fda/fdacee/02-09.html>
- de la Fuente, A. (2009). "Un enlace alternativo de los agregados de VAB y empleo de la CRE95 y la CRE00." Mimeo, Instituto de Análisis Económico, CSIC, Barcelona.  
[http://www.fedea.es/033\\_Publicaciones\\_VerMas.asp?id=841](http://www.fedea.es/033_Publicaciones_VerMas.asp?id=841)
- de la Fuente, A. y R. Doménech (2006). "Human capital in growth regressions: how much difference does data quality make?" *Journal of the European Economic Association* 4(1), pp. 1-36.
- Doménech, R., J. Escribá y M. J. Murgui (1999). "Cambios en precios relativos y crecimiento económico en las regiones españolas." *Revista de Economía Aplicada* VII, pp. 5-29.
- Fundación BBV (FBBV, 1999). *Renta nacional de España y su distribución provincial. Serie homogénea. Años 1955 a 1993 y avances 1994 a 1997*. Bilbao.
- Fundación BBVA (FBBVA, 2000). *Renta nacional de España y su distribución provincial. Año 1995 y avances 1996-1999*. Bilbao.
- Instituto Nacional de Estadística (INE, 1961). *Anuario Estadístico de España, Año 1961*. En base de datos INEbase Historia. Madrid. <http://www.ine.es/inebaseweb/25687.do>
- Instituto Nacional de Estadística (INE, 1962). *Censo de la Población y de las Viviendas de España según la inscripción realizada el 31 de diciembre de 1960*. En base de datos INEbase: Demografía y población. Cifras de población y censos demográficos. Censos de población desde 1900.  
<http://www.ine.es/inebaseweb/71807.do?language=0>
- Instituto Nacional de Estadística (INE, 2005a). "Cambio de base de la Contabilidad Nacional de España." 19 de mayo de 2005.
- Instituto Nacional de Estadística (INE, 2005b). "Contabilidad Nacional de España (CNE). Base 2000. Nota informativa." 19 de mayo de 2005.
- Instituto Nacional de Estadística (INE, 2007). "Contabilidad regional de España, base 2000. Serie homogénea 1995-2006. Nota metodológica." En Base de datos electrónica INEbase. Economía: Cuentas Económicas: Contabilidad Regional de España. Madrid.  
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft35%2Fp010&file=inebase&L=0>

- Instituto Nacional de Estadística (INE, 2009a). Contabilidad Nacional de España. Datos históricos, Base 1986. En Base de datos electrónica INEbase. Economía. Cuentas Económicas. Contabilidad Nacional de España. Madrid.  
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t35/p008/cne/ba86&file=pcaxis>
- Instituto Nacional de Estadística (INE, 2009b). Contabilidad Regional de España, Base 1986. En Base de datos electrónica INEbase. Economía: Cuentas Económicas. Contabilidad Regional de España. Madrid.  
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t35/p010/a1996&file=pcaxis>
- Instituto Nacional de Estadística (INE, 2009c). Cifras de población. En Base de datos electrónica INEbase. Demografía y población. Madrid.  
[http://www.ine.es/inebmenu/mnu\\_cifraspob.htm](http://www.ine.es/inebmenu/mnu_cifraspob.htm)
- Instituto Nacional de Estadística (INE, 2008a). Contabilidad Regional de España. En Base de datos electrónica INEbase. Economía: Cuentas Económicas. Madrid.  
<http://www.ine.es/inebase/cgi/um?M=%2Ft35%2Fp010&O=inebase&N=&L=>
- Krueger, A. y M. Lindahl (2001). "Education for growth: why and for whom?" *Journal of Economic Literature* XXXIX, pp. 1101-36.
- Maluquer, J. (2006?). "El desarrollo económico de las Islas Baleares en perspectiva comparada (1950-2000)." Mimeo, Universidad Autónoma de Barcelona.
- Maluquer, J. (2008). "El crecimiento moderno de la población en España de 1850 a 2001: una serie homogénea anual." *Investigaciones de Historia Económica* 10, pp. 129-62.
- Maluquer, J. y M. Llonch (2005). "Trabajo y relaciones laborales." En A. Carreras y X. Tafunell, coordinadores. *Estadísticas históricas de España, siglos XIX-XX*, segunda edición. Fundación BBVA, Bilbao, pp. 1155-1245.
- Martin Rodriguez, M. (1992). "Pautas y tendencias de desarrollo económico regional en España: una visión retrospectiva." En J. L. García Delgado y A. Pedreño, directores, *Ejes territoriales de desarrollo: España en la Europa de los noventa*. Madrid, Economistas, pp. 133-55.
- Martín Rodriguez, M. (1993). "Evolución de las disparidades económicas regionales: una perspectiva histórica." En J. L. García Delgado, editor, *España, Economía*. Espasa Calpe, Madrid, pp. 891-927.
- Plaza Prieto, J. (1953). "El producto nacional de España y su distribución espacial: un análisis hipotético." *De Economía* VI(22), pp. 9-31. Reproducido en J. Velarde, editor (1969). *Lecturas de economía española*. Madrid, Gredos, pp. 198-218.
- Uriel, E., M. L. Moltó y V. Cucarella (2000). "Contabilidad Nacional de España. Series enlazadas 1954- 1997. (CNEe-86). Fundación BBV, Bilbao.

## ANEXO 1

### 1. ¿Qué serie de referencia? Detalles técnicos

En este apartado se presentan los detalles técnicos del procedimiento utilizado en el texto para fijar la fecha de corte en la que se abandona la serie de la Fundación BBVA para pasar a tomar como referencia las series de la CRE86. Los apartados 1.1 a 1.4 recogen algunos resultados y desarrollos preliminares. Este material se utiliza en los apartados 1.5 y 1.6 para calcular el error cuadrático medio (ECM) de las distintas series, el nivel crítico del parámetro  $\phi$  que mide el posible sesgo hacia cero de las series del BBVA y la cota superior para el mismo parámetro.

#### 1.1. Estimación por MCO cuando no hay errores de medición

Como referencia, conviene repasar algunos resultados básicos sobre el modelo estándar de regresión que utilizaré como punto de partida en lo que sigue. Supongamos que  $y$  viene dada por

$$(1) y_i = \beta x_i + u_i$$

donde  $u_i$  es una perturbación aleatoria iid con media cero y varianza  $\sigma_u^2$ . Supongamos también que se cumplen las hipótesis del modelo clásico de regresión y, en particular, que el término de error no está correlacionado con el regresor ( $Ex_i u_i = 0$ ) de forma que la estimación del modelo por mínimos cuadrados ordinarios (MCO) producirá estimadores consistentes del parámetro  $\beta$ . Para simplificar los cálculos he supuesto también (sin pérdida de generalidad) que todas las variables tienen media cero, lo que entre otras cosas implica que la constante de la regresión desaparece.

La ecuación (1) implica que la varianza de  $y_i$  dada  $x_i$  es igual a

$$\text{var}(y_i | x_i) = \sigma_u^2 \quad \text{para todo } i$$

donde conviene observar que la varianza “condicional”  $\text{var}(y_i | x_i)$  no es lo mismo que la varianza “total” o incondicional de  $y$ ,  $\text{var } y_i$ .

El estimador MCO de  $\beta$  viene dado por

$$(2) b = \frac{\sum x_i y_i}{\sum x_i^2} = \sum \left( \frac{x_i}{\sum x_i^2} \right) y_i$$

y, por tanto, su varianza (condicional en  $x$ ) será igual a

$$(3) \text{var } b = \sum \left( \frac{x_i}{\sum x_i^2} \right)^2 \text{var}(y_i | x_i) = \sigma_u^2 * \sum \left( \frac{x_i}{\sum x_i^2} \right)^2 = \sigma_u^2 * \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{\sigma_u^2}{\sum x_i^2}$$

Por último, el  $R^2$  de la regresión es la fracción de la variación total de  $y$  que viene explicada por el modelo. Utilizando (2), tenemos

$$(4) R^2(y|x) = \frac{\sum (bx_i)^2}{\sum y_i^2} = \frac{b^2 \sum x_i^2}{\sum y_i^2} = \frac{\frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \sum x_i^2}{\sum y_i^2} = \frac{(\sum x_i y_i)^2}{(\sum x_i^2)(\sum y_i^2)}$$

## 1.2. Implicaciones del error de medición en la variable dependiente

Consideremos ahora la posibilidad de que la variable dependiente,  $y$ , esté medida con error y pueda presentar un sesgo sistemático hacia cero. Supondremos que lo que podemos observar no es  $y$ , sino una o varias *proxies* ruidosas de la forma

$$(5) y_i^k = (1 - \phi_k)y_i + e_i^k$$

donde  $\phi_k \in [0,1]$  captura el posible sesgo sistemático en la medición de  $y$  y  $e_i^k$  es un error “clásico” de medición iid y con

$$(6) Ee_i^k = 0, E(e_i^k)^2 = \sigma_k^2 \text{ para todo } i \text{ y } Ee_i^k x_i = Ee_i^k y_i = 0$$

Observemos que

$$(7) \text{var}(y_i^k | x_i) = (1 - \phi_k)^2 \text{var}(y_i | x_i) + \text{var } e_i^k = (1 - \phi_k)^2 \sigma_u^2 + \sigma_e^2$$

Supongamos ahora que estimamos la regresión dada en (1) utilizando  $y^k$  en vez de  $y$  como variable dependiente. El estimador de MCO vendrá dado por

$$(8) b_k = \frac{\sum x_i y_i^k}{\sum x_i^2} = \sum \left( \frac{x_i}{\sum x_i^2} \right) y_i^k$$

y su varianza será

$$(9) \text{var } b_k = \frac{\text{var}(y_i^k | x_i)}{\sum x_i^2} = \frac{(1 - \phi_k)^2 \sigma_u^2 + \sigma_e^2}{\sum x_i^2}$$

Es fácil comprobar que el sesgo sistemático en la medición de  $y$  se traslada a  $b_k$ . Usando (8) y (5) tenemos

$$(10) b_k = \frac{\sum x_i y_i^k}{\sum x_i^2} = \frac{\sum x_i [(1 - \phi_k)y_i + e_i^k]}{\sum x_i^2} = \frac{(1 - \phi_k) \sum x_i y_i}{\sum x_i^2} + \frac{\sum x_i e_i^k}{\sum x_i^2} = (1 - \phi_k)b + \frac{\sum x_i e_i^k}{\sum x_i^2}$$

donde  $b$  es el estimador de MCO que se obtendría si  $y$  fuese directamente observable. Puesto que  $b$  es consistente y por hipótesis  $Ee_i^k x_i = 0$ , tenemos

$$(11) Eb_k = (1 - \phi_k)b$$

Examinando (9) y (11), vemos que el componente clásico del error de medición en  $y$  no sesga el estimador pero aumenta su varianza. El componente sistemático, por el contrario, genera un sesgo a la baja en el estimador pero también reduce su varianza.

Seguidamente, queremos calcular el error cuadrático medio (ECM) del estimador MCO de  $\beta$ . Definiendo el sesgo asintótico de  $b_k$  como

$$(12) \quad s_k = Eb_k - \beta = (1 - \phi_k)\beta - \beta = -\phi_k\beta$$

tenemos

$$(13) \quad \begin{aligned} E(b_k - \beta)^2 &= E((b_k - Eb_k) + (Eb_k - \beta))^2 = E((b_k - Eb_k) + s_k)^2 \\ &= E((b_k - Eb_k)^2 + 2(b_k - Eb_k)s_k + s_k^2) = \\ &= E(b_k - Eb_k)^2 + 2s_k E(b_k - Eb_k) + s_k^2 = E(b_k - Eb_k)^2 + 0 + s_k^2 \\ &= \text{var } b_k + \phi_k^2 \beta^2 \end{aligned}$$

Substituyendo (9) en esta expresión y dividiendo por  $\beta^2$  tenemos

$$(14) \quad E_k \equiv \frac{E(b_k - \beta)^2}{\beta^2} = \phi_k^2 + \frac{\text{var } b_k}{\beta^2} = \phi_k^2 + \frac{(1 - \phi_k)^2 \sigma_u^2 + \sigma_k^2}{\beta^2 \sum x_i^2}$$

Nuestro objetivo es estimar el valor de  $E_k$  para las dos series alternativas con las que estamos trabajando (la serie del BBVA y la de CRE) con el fin de determinar cuál de ellas es “mejor” en el sentido de generar estimadores con menor ECM en cada uno de los años en los que ambas se solapan. Más adelante volveré a esta expresión e intentaré expresarla en términos de magnitudes observables o estimables (que a poder ser no dependan de las  $x$ 's, dado que  $x$  puede ser en principio cualquier variable). Antes, sin embargo, necesitamos introducir algunos conceptos y realizar algunos cálculos preliminares.

### 1.3. Ratios de fiabilidad

Supongamos que disponemos de dos *proxies* alternativas para  $y$ ,  $y^j$  e  $y^k$ . Regresando cada una de estas variables sobre la otra podemos obtener un estimador del ratio entre la varianza de la variable subyacente,  $y$ , y la varianzas de cada una de las *proxies*. Este cociente se suele denominar *ratio de fiabilidad* y, bajo ciertas condiciones, constituye un buen indicador del contenido informativo de las distintas series (esto es, un indicador inverso de la importancia del “ruido” debido a la existencia de errores de medición).<sup>1</sup>

Para concretar, identificaremos  $y^k$  con la serie del BBVA e  $y^j$  con la serie enlazada de la CRE. Esto es, supondremos que la segunda serie no presenta un sesgo sistemático ( $\phi_j = 0$ ) pero a cambio podría contener más ruido puro ( $\sigma_j^2 > \sigma_k^2$ ). Tenemos, por tanto,

$$(15) \quad y_i^k = (1 - \phi_k)y_i + e_i^k$$

$$(16) \quad y_i^j = y_i + e_i^j$$

---

<sup>1</sup> Véase por ejemplo Krueger y Lindahl (2001) y de la Fuente y Doménech (2006).

donde supondremos que los errores  $e_i^k$  y  $e_i^j$ , además de no estar correlacionados con  $x$  e  $y$ , tampoco lo están entre sí ( $Ee_i^j e_i^k = 0$ ). Utilizando estas hipótesis, tenemos

$$(17) E y_i^k y_i^j = E[(1 - \phi_k) y_i + e_i^k][y_i + e_i^j] = (1 - \phi_k) E y_i^2$$

Consideremos ahora una regresión de la forma  $y_i^j = r_k y_i^k + \varepsilon_{ki}$ . El valor esperado del estimador del coeficiente de pendiente (el ratio de fiabilidad de  $y_i^k$ ) viene dado por

$$\hat{r}_k = \frac{\sum y_i^k y_i^j}{\sum (y_i^k)^2} = (1 - \phi_k) \frac{\sum y_i^2}{\sum (y_i^k)^2} + (1 - \phi_k) \frac{\sum y_i e_i^j}{\sum (y_i^k)^2} + (1 - \phi_k) \frac{\sum y_i e_i^k}{\sum (y_i^k)^2} + \frac{\sum e_i^k e_i^j}{\sum (y_i^k)^2}$$

Esta expresión es difícil de simplificar porque contiene cocientes de variables aleatorias (obsérvese que a diferencia de  $x$ , no podemos tratar a  $y_i^k$  como una constante). Sin embargo, sí podemos calcular el límite en probabilidad del estimador, que viene dado por

$$(18) \tilde{E} \hat{r}_k \equiv \text{plim } \hat{r}_k = \frac{E y_i^k y_i^j}{E (y_i^k)^2} = \frac{(1 - \phi_k) E y_i^2}{E (y_i^k)^2}$$

Utilizando esta expresión, tenemos

$$(19) \frac{E y_i^2}{E (y_i^k)^2} = \frac{\tilde{E} \hat{r}_k}{(1 - \phi_k)}$$

Intercambiando los papeles de las dos series, tenemos también

$$(20) \tilde{E} \hat{r}_j = \frac{E y_i^k y_i^j}{E (y_i^j)^2} = \frac{(1 - \phi_k) E y_i^2}{E (y_i^j)^2} \Rightarrow \frac{E y_i^2}{E (y_i^j)^2} = \frac{\tilde{E} \hat{r}_j}{(1 - \phi_k)}$$

Más adelante, utilizaremos estas expresiones para escribir el ECM de cada serie en términos de magnitudes “observables.”

#### 1.4. Otros preliminares

La expresión que hemos derivado para el ECM del estimador de  $\beta$  en presencia de errores de medición (ecuación (14)) incluye las varianzas de la perturbación de la regresión original y del componente clásico del error de medición ( $\sigma_u^2$  y  $\sigma_k^2$ ). Queremos reemplazar estos dos términos por otras magnitudes más fáciles de estimar o aproximar sin usar las  $x$ 's.

Utilizando

$$(1) y_i = \beta x_i + u_i$$

podemos calcular la varianza “total” (no condicional) de  $y$ :

$$(21) \text{var } y_i = \beta^2 \text{var } x_i + \sigma_u^2$$

y escribir  $\sigma_u^2$  en la forma



$$(22) \sigma_u^2 = \text{var } y_i - \beta^2 \text{var } x_i$$

De la misma forma, dada

$$(15) y_i^k = (1 - \phi_k) y_i + e_i^k$$

tenemos

$$(23) \text{var } y_i^k = (1 - \phi_k)^2 \text{var } y_i + \sigma_k^2 \Rightarrow \sigma_k^2 = \text{var } y_i^k - (1 - \phi_k)^2 \text{var } y_i$$

Substituyendo (22) y (23) en (14), tenemos:

$$\begin{aligned} E_k &= \phi_k^2 + \frac{(1 - \phi_k)^2 \sigma_u^2 + \sigma_k^2}{\beta^2 \sum x_i^2} = \phi_k^2 + \frac{(1 - \phi_k)^2 [\text{var } y_i - \beta^2 \text{var } x_i] + \text{var } y_i^k - (1 - \phi_k)^2 \text{var } y_i}{\beta^2 \sum x_i^2} \\ (24) \quad &= \phi_k^2 + \frac{-(1 - \phi_k)^2 \beta^2 \text{var } x_i + \text{var } y_i^k}{\beta^2 \sum x_i^2} = \phi_k^2 - \frac{(1 - \phi_k)^2 \text{var } x_i}{\sum x_i^2} + \frac{\text{var } y_i^k}{\beta^2 \sum x_i^2} \\ &= \phi_k^2 - \frac{(1 - \phi_k)^2 \text{var } x_i}{\sum x_i^2} + \frac{\text{var } y_i^k}{\text{var } y_i} \frac{\text{var } y_i}{\beta^2 \sum x_i^2} \end{aligned}$$

### 1.5. Estimación del ECM

Seguidamente, construimos un estimador de  $E_k$  reemplazando los momentos poblacionales que aparecen en (24) por sus homólogos muestrales y los parámetros (excepto  $\phi_k$  que quedará como un parámetro “libre”) por estimadores insesgados o al menos consistentes. En particular, reemplazamos

$$\text{var } x_i \quad \text{por} \quad \frac{1}{n-1} \sum x_i^2 \quad \text{donde } n \text{ es el número de observaciones}$$

$$\frac{\text{var } y_i^k}{\text{var } y_i} \quad \text{por} \quad \frac{1 - \phi_k}{\hat{r}_k} \quad \text{véase la ecuación (19)}$$

$$\beta^2 \quad \text{por} \quad b^2 = \frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \quad \text{véase la ecuación (2) y}$$

$$\text{var } y_i \quad \text{por} \quad \frac{1}{n-1} \sum y_i^2 \quad \text{y}$$

$$\tilde{E} \hat{r}_k \quad \text{por} \quad \hat{r}_k$$

para obtener un estimador consistente de  $E_k$

$$\begin{aligned}
\hat{E}_k &= \phi_k^2 - \frac{(1-\phi_k)^2}{n-1} \frac{\sum x_i^2}{\sum x_i^2} + \frac{1-\phi_k}{\hat{r}_k} \frac{\frac{1}{n-1} \sum y_i^2}{\frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \sum x_i^2} \\
(25) \quad &= \phi_k^2 - \frac{(1-\phi_k)^2}{n-1} + \frac{1-\phi_k}{(n-1)\hat{r}_k} \frac{(\sum y_i^2)(\sum x_i^2)}{(\sum x_i y_i)^2} = \phi_k^2 - \frac{(1-\phi_k)^2}{n-1} + \frac{1-\phi_k}{(n-1)\hat{r}_k R^2(x)} \\
&= \phi_k^2 - \frac{1}{n-1} \left( (1-\phi_k)^2 - \frac{1-\phi_k}{\hat{r}_k R^2(x)} \right)
\end{aligned}$$

donde hemos hecho uso de la ecuación (4) para escribir el último término de la expresión en función del  $R^2$  de una regresión de  $y$  sobre  $x$ ,  $R^2(x)$ . Obsérvese que el último término tiende a cero cuando  $n \rightarrow \infty$ . Esto es, el ECM se reduce al cuadrado del sesgo cuando el tamaño de la muestra tiende a infinito.

Procediendo de la misma forma en el caso de  $y^j$ , recordando que por hipótesis  $\phi_j = 0$  y utilizando la ecuación (20), tenemos:

$$(26) \quad \hat{E}_j = -\frac{1}{n-1} \frac{\sum x_i^2}{\sum x_i^2} + \frac{1-\phi_k}{\hat{r}_j} \frac{\frac{1}{n-1} \sum y_i^2}{\frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \sum x_i^2} = -\frac{1}{n-1} + \frac{1-\phi_k}{(n-1)\hat{r}_j R^2(x)} = \frac{1}{n-1} \left( \frac{1-\phi_k}{\hat{r}_j R^2(x)} - 1 \right)$$

Definamos la función

$$(27) \quad F(\phi_k, R^2(x)) = (n-1)(\hat{E}_k - \hat{E}_j)$$

Diremos que la serie  $y^k$  es “mejor” que la serie  $y^j$  si tiene un ECM menor, esto es, si  $F(\phi_k, R^2(x)) < 0$ . Puesto que  $F()$  es una función de  $x$ , por el momento la afirmación de que una serie es mejor que otra dependerá de la variable  $x$  que estemos considerando y, por supuesto, del valor del parámetro  $\phi_k$ , sobre el que no sabemos nada.

Combinando (25) y (26) tenemos

$$\begin{aligned}
(28) \quad F(\phi_k, R^2(x)) &= (n-1)(\hat{E}_k - \hat{E}_j) = (n-1)\phi_k^2 - (1-\phi_k)^2 + \frac{1-\phi_k}{\hat{r}_k R^2(x)} - \frac{1-\phi_k}{\hat{r}_j R^2(x)} + 1 \\
&= (n-1)\phi_k^2 - (1-2\phi_k + \phi_k^2) + \frac{1-\phi_k}{R^2(x)} \left( \frac{1}{\hat{r}_k} - \frac{1}{\hat{r}_j} \right) + 1 \\
&= (n-1)\phi_k^2 + (2\phi_k - \phi_k^2) + \frac{1-\phi_k}{R^2(x)} \left( \frac{\hat{r}_j - \hat{r}_k}{\hat{r}_k \hat{r}_j} \right) \\
&= (n-2)\phi_k^2 + 2\phi_k - \frac{1-\phi_k}{R^2(x)} \left( \frac{\hat{r}_k - \hat{r}_j}{\hat{r}_k \hat{r}_j} \right) \\
&= (n-2)\phi_k^2 + \left( 2 + \frac{1}{R^2(x)} \left( \frac{\hat{r}_k - \hat{r}_j}{\hat{r}_k \hat{r}_j} \right) \right) \phi_k - \frac{1}{R^2(x)} \left( \frac{\hat{r}_k - \hat{r}_j}{\hat{r}_k \hat{r}_j} \right) \\
&\equiv (n-2)\phi_k^2 + \left( 2 + \frac{A}{R^2(x)} \right) \phi_k - \frac{A}{R^2(x)}
\end{aligned}$$

que es la ecuación (4) en el texto.

### 1.6. Una cota para $\phi$

El ratio de fiabilidad de  $y_i^j$  puede utilizarse para estimar una cota superior para  $\phi_k$ . Dado que

$$(16) \quad y_i^j = y_i + e_i^j$$

tenemos

$$E(y_i^j)^2 = Ey_i^2 + E(e_i^j)^2 = Ey_i^2 + \sigma_j^2$$

Dividiendo por  $Ey_i^2$

$$\frac{E(y_i^j)^2}{Ey_i^2} = 1 + \frac{\sigma_j^2}{Ey_i^2}$$

Utilizando

$$(20) \quad \tilde{E}r_j = \frac{Ey_i^k y_i^j}{E(y_i^j)^2} = \frac{(1 - \phi_k)Ey_i^2}{E(y_i^j)^2} \Rightarrow \frac{Ey_i^2}{E(y_i^j)^2} = \frac{\tilde{E}r_j}{(1 - \phi_k)}$$

en la expresión anterior

$$\frac{(1 - \phi_k)}{\tilde{E}r_j} = 1 + \frac{\sigma_j^2}{Ey_i^2} \Rightarrow (1 - \phi_k) = \tilde{E}r_j \left( 1 + \frac{\sigma_j^2}{Ey_i^2} \right) \geq \tilde{E}r_j$$

de donde

$$\phi_k \leq 1 - \tilde{E}r_j$$

Puesto que  $\hat{r}_j$  es (por definición) un estimador consistente de  $\tilde{E}r_j$ , podemos utilizarlo para construir una cota tentativa para  $\phi_k$ . Si además suponemos que este último parámetro es constante durante el período que estamos analizando (lo que parece razonable, dado que la metodología del BBVA ya estaba bien asentada en 1981 y más aún en el momento en que se revisan las series para construir la serie homogénea 1955-95), tendremos

$$\phi_k \leq 1 - \tilde{E}r_{jt} \quad \text{para todo } t$$

y por lo tanto

$$\phi_k \leq 1 - \max_t \tilde{E}r_{jt} \equiv \phi_{\max}$$

lo que nos permite estimar  $\phi_{\max}$  como

$$(21) \quad \hat{\phi}_{\max} = 1 - \max_t \hat{r}_{jt}$$

## 2. Enlace directo de magnitudes reales

En lo que sigue, utilizaré la siguiente notación. Sean  $XN_t$  e  $YN_t$  las series a enlazar, medidas en términos "nominales" (esto es, a precios corrientes) y  $XR_t^{BX}$  e  $YR_t^{BY}$  las correspondientes magnitudes "reales" medidas a precios del año base de cada serie,  $BX$  y  $BY$ . Los correspondientes índices de precios con la misma base son los cocientes entre las series nominales y reales,

$$(22) \quad XP_t^{BX} = \frac{XN_t}{XR_t^{BX}} \quad \text{y} \quad YP_t^{BY} = \frac{YN_t}{YR_t^{BY}}$$

o, utilizando minúsculas para indicar que estamos trabajando con logaritmos,

$$(22') \quad xp_t^{BX} = xn_t - xr_t^{BX} \quad \text{y} \quad yp_t^{BY} = yn_t - yr_t^{BY}$$

El enlace de las series nominales por el procedimiento mixto vendrá dado por

$$(23) \quad \widehat{yn}_t = xn_t + \widehat{dn}_t \quad \text{para } t \leq T \quad \text{con} \quad \widehat{dn}_t = dn_T \rho_n^{\frac{T-t}{T}} = (yn_T - xn_T) \rho_n^{\frac{T-t}{T}}$$

El primer paso para enlazar las series reales consiste en expresar  $X$  e  $Y$  a precios del año de enlace,  $T$ . Para ello, tenemos que modificar el valor de la serie expresado en precios del año base original en proporción al incremento de precios observado entre ese año y el año de enlace, esto es, construir

$$(24) \quad XR_t^{BT} = XR_t^{BX} \frac{XP_T^{BX}}{XP_T^{BX}} = XR_t^{BX} XP_T^{BX} \quad \text{y} \quad YR_t^{BT} = YR_t^{BY} YP_T^{BY}$$

donde se ha hecho uso del hecho de que el índice de precios en el año base es igual a la unidad por definición. En logaritmos, tendríamos

$$(24') \quad xr_t^{BT} = xr_t^{BX} + xp_T^{BX} \quad \text{y} \quad yr_t^{BT} = yr_t^{BY} + yp_T^{BY}$$

Seguidamente enlazamos la serie a precios del año de enlace de forma que

$$\widehat{yr}_t^{BT} = xr_t^{BT} + \widehat{dr}_t^{BT} \quad \text{para } t \leq T \quad \text{con} \quad \widehat{dr}_t^{BT} = dr_T^{BT} \rho_r^{\frac{T-t}{T}} = (yr_T^{BT} - xr_T^{BT}) \rho_r^{\frac{T-t}{T}}$$

Observando que las magnitudes reales y nominales coinciden por definición en el año base, tenemos

$$(25) \quad \widehat{yr}_t^{BT} = xr_t^{BT} + (yr_T^{BT} - xr_T^{BT}) \rho_r^{\frac{T-t}{T}} = xr_t^{BT} + (yn_T - xn_T) \rho_r^{\frac{T-t}{T}}$$

Finalmente, utilizamos (24') para recuperar la serie enlazada a precios del año base original de la serie más reciente,  $BY$ ,

$$(26) \quad \widehat{yr}_t^{BY} = \widehat{yr}_t^{BT} - yp_T^{BY} = xr_t^{BT} + (yn_T - xn_T) \rho_r^{\frac{T-t}{T}} - yp_T^{BY}$$

Utilizando (22') y (24'), tenemos,

$$x_t^{BT} = x_t^{BX} + xp_t^{BX} = xn_t - xp_t^{BX} + xp_t^{BX}$$

lo que implica que (26) se puede escribir de la forma

$$(27) \quad \widehat{y}_t^{BY} = x_t^{BT} + (y_{n_T} - xn_T) \rho_r^{\frac{T-t}{T}} - yp_T^{BY} = xn_t - xp_t^{BX} + (y_{n_T} - xn_T) \rho_r^{\frac{T-t}{T}} + xp_t^{BX} - yp_T^{BY}$$

Finalmente, la serie enlazada de índices de precios en base *BY* vendrá dada por el cociente entre la serie enlazada a precios corrientes y la serie enlazada a precios de *BY*:

$$(28) \quad \widehat{y}_t^{BY} = \widehat{y}_t - \widehat{y}_t^{BY} = xp_t^{BX} + (yp_T^{BY} - xp_T^{BX}) + (y_{n_T} - xn_T) \left( \rho_n^{\frac{T-t}{T}} - \rho_r^{\frac{T-t}{T}} \right)$$

Obsérvese esta expresión se reduce a la retropolación de las series originales de precios cuando el último término desaparece, esto es, cuando  $\rho_n = \rho_r$ . Por consiguiente, para garantizar el patrón deseado de enlace de las series de deflatores implícitos, el enlace de la serie de volumen o precios constantes ha de realizarse utilizando el mismo valor de  $\rho$  que para la serie a precios corrientes.

### 3. Construcción de las series de referencia a nivel nacional para 1955-95

#### 3.1. VAB a coste de los factores

Uriel, Moltó y Cucarella (2000) construyen series homogéneas de Contabilidad Nacional en base 1986 para el período 1954-1997 extendiendo hacia atrás la CNE base 86 (CNE86) mediante su enlace por retropolación con series anteriores, apoyándose en parte en trabajos anteriores del propio INE. El trabajo de Uriel et al, junto con la propia CNE86 y la CRE86, se utiliza como punto de partida para construir una serie enlazada y corregida de VAB a coste de los factores (VABcf) a precios corrientes y constantes de 1986 que posteriormente se enlazarán con series posteriores en base 2000 para obtener las series definitivas de VAB a nivel nacional (excluyendo la “extra-regio”).

Siempre que ha sido posible, he tomado los datos directamente de las series de CNE86 o CRE86 que ofrece el INE en su página web (INE 2009a,b). Las series ofrecidas en estas fuentes comienzan en 1971 o en 1986, dependiendo de la magnitud. Para años anteriores, los datos se toman de Uriel, Moltó y Cucarella (2000). He comprobado que las dos fuentes coinciden (con discrepancias muy menores en algún caso) durante su período de solapamiento.

De 1986 en adelante, la CRE86 ofrece datos directos de VABcf a precios corrientes para España en su conjunto y para cada uno de sus territorios, incluyendo la “extra-regio.” Para años anteriores, el VABcf se recupera a partir del PIB o VAB a precios de mercado y de las series de impuestos indirectos y de subvenciones de explotación. La relación entre los diversos agregados de renta, impuestos y subvenciones se recoge en el Cuadro A.1.1, ilustrándose con datos de 1986. Para los años anteriores a 1986, la serie de IVA se extiende hacia atrás en la llamada “línea fiscal homogénea” como la suma de los rendimientos de los impuestos

indirectos a los que vino a sustituir el IVA (entre otros, el impuesto general sobre el tráfico de empresas y el impuesto sobre el lujo).

**Cuadro A1.1: Relación entre distintos agregados en la CNE86, año 1986**

	<i>Meuros</i>	<i>% del total</i>
<i>PIB a precios de mercado</i>	194,271	100.0%
- IVA y antecesores (línea fiscal homogénea)	-8,822	-4.5%
- impuestos ligados a la importación, excluido el IVA y antecesores	-1,978	-1.0%
+ subvenciones a la importación	0.2	0.0%
= <i>VAB a precios de mercado</i>	183,471	94.4%
- impuestos ligados a la producción, excluido el IVA y antecesores	-10,178	-5.2%
+ subvenciones de explotación	4,043	2.1%
= <i>VAB a coste de los factores</i>	177,336	91.3%

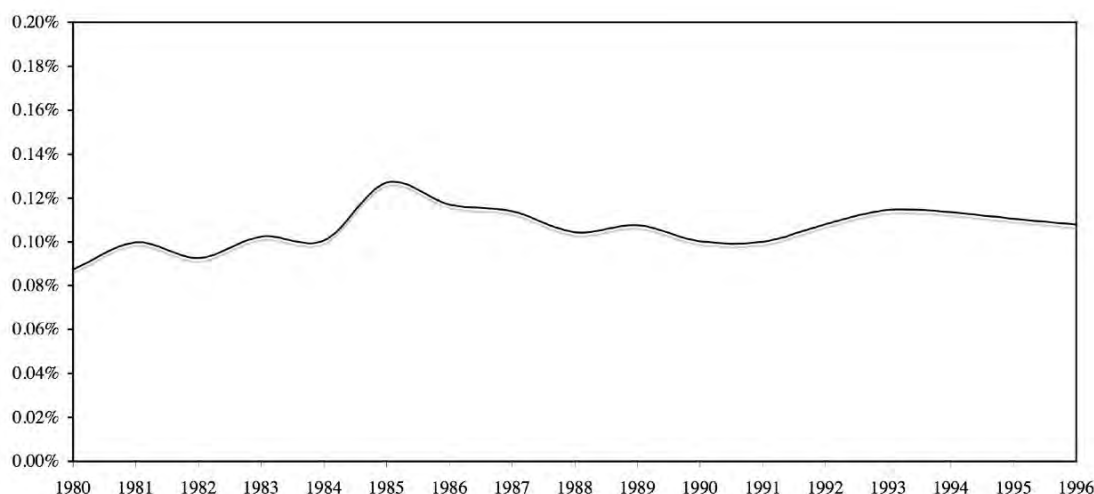
- Fuente: CRE86 y CNE86.

Ninguna de las fuentes citadas contiene la información necesaria para construir series de VABcf a precios constantes. Sin embargo, sí es posible construir una serie de VAB a precios de mercado (VABpm) a precios constantes que, confrontada con la correspondiente serie a precios corrientes, nos permite obtener un deflactor para esta magnitud. Este índice de precios es el que he utilizado para deflactar la serie de VABcf a precios corrientes, obteniendo así una serie aproximada de VABcf a precios constantes de 1986.

#### *Correcciones a la serie de VABcf*

En la serie cuya construcción acabo de describir se introducen dos correcciones con el fin de hacerla lo más homogénea posible con la serie posterior con la que se terminará enlazando. Por un lado, se subtrae del agregado nacional una estimación del VABcf que corresponde a la llamada “extra-regio” y por otro se añade al mismo agregado una estimación de la fracción de la llamada producción imputada de servicios bancarios (PISB, posteriormente SIFMI) que se habría incluido en el VAB con las normas actuales de Contabilidad Nacional y Regional.

**Gráfico A1.1: Peso de la “extra-regio” en el VABcf total de España**



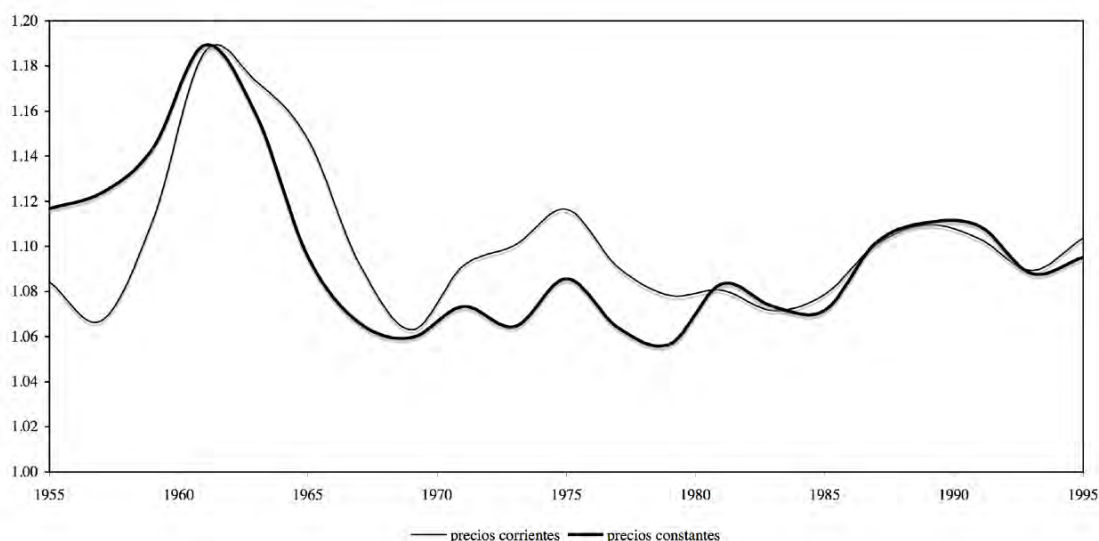
El Gráfico A1.1 muestra que el peso de la “extra-regio” en el VABcf agregado se mantiene aproximadamente constante durante el período para el que se dispone de datos (tomados de la CRE86 completada hasta 1980 como se indica en la sección 5 del texto), aunque con ligeras oscilaciones y quizás con una ligera tendencia al alza. Para corregir la serie en años anteriores a 1980, supondré un peso de la extra-regio en el VAB agregado igual al observado en el año citado (el 0.0875%). Este porcentaje se aplicará al VAB original, tanto a precios corrientes como a precios constantes, para realizar la corrección necesaria.

En cuanto a la PISB, con ocasión del cambio metodológico que se produce en la base 2000 en relación con el tratamiento de los SIFMI, el INE (2005b) ofrece un desglose de esta magnitud por sectores institucionales en el año 2000 del que se desprende que un 39.41% de este agregado deja de considerarse parte del consumo intermedio para integrarse en el consumo final y por lo tanto en el VAB. A falta de otra información, supondré que este porcentaje se ha mantenido constante en el tiempo y lo utilizaré para estimar la parte de la PISB (a precios corrientes y constantes) que ha de sumarse al VABcf en cada año.

#### *Comparación con las series del BBVA*

El Gráfico A1.2 revela que existen diferencias muy significativas entre las series de VAB construidas en el apartado anterior a partir de la Contabilidad Nacional y las elaboradas por Julio Alcaide y sus colaboradores para la Fundación BBVA. La segunda serie se sitúa consistentemente por encima de la primera, con una diferencia que se aproxima al 20% a comienzos de la década de los sesenta tanto en términos nominales como reales. La diferencia se estrecha posteriormente, pero se mantiene en torno al 10% hasta los años finales de la muestra. En algunos períodos, la diferencia entre las dos series dibujadas en el gráfico es significativa, lo que indica una cierta divergencia entre los índices de precios utilizados en ambas fuentes. Esta diferencia, sin embargo, prácticamente desaparece en la segunda mitad de la muestra.

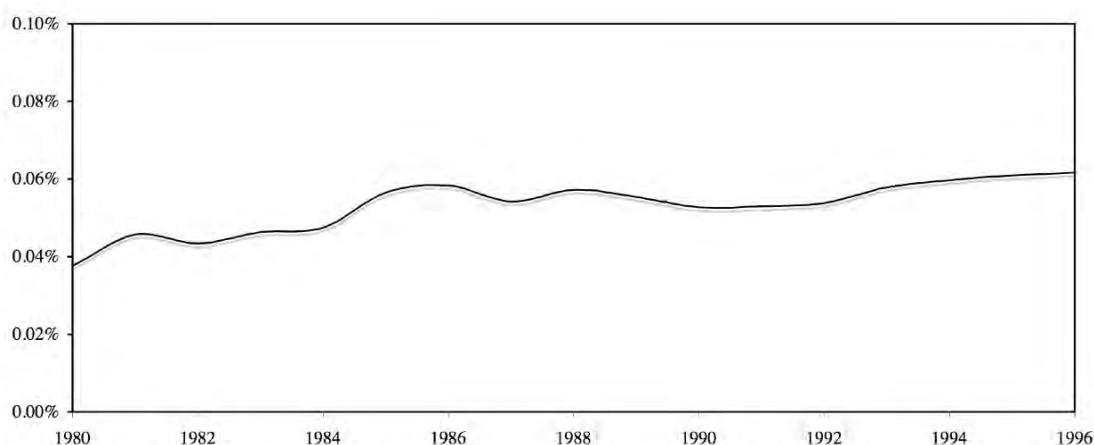
**Gráfico A1.2: VABcf a precios corrientes y constantes, ratio FBBVA/CNEe86**



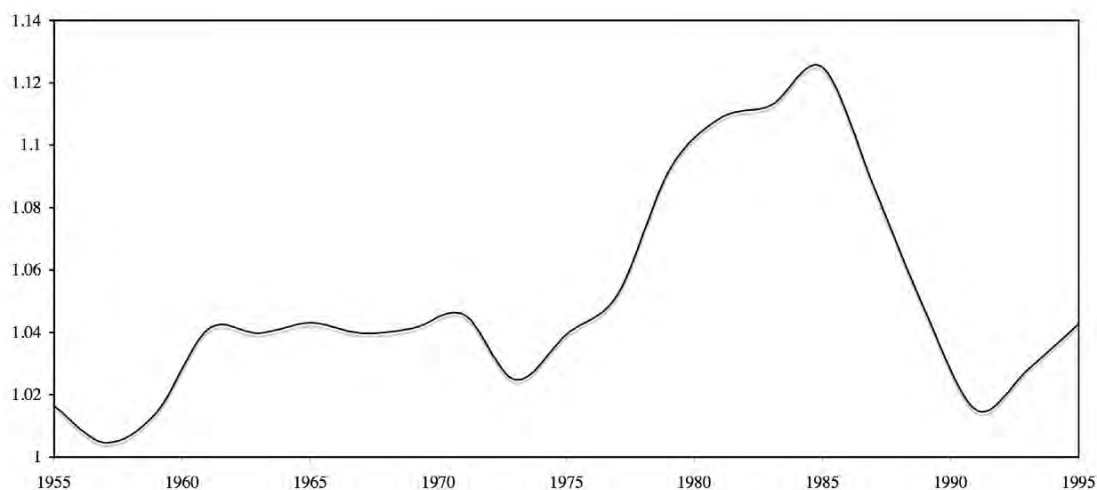
### 3.2. Empleo

En INE (2009a) se ofrecen series de empleo (medido por el número de ocupados) para el período 1971-97. Maluquer y Llonch (M&L, 2000) extienden esta serie, enlazándola con versiones anteriores y posteriores de la CNE. La serie de M&L presenta una clara discontinuidad en 1995, pues en ese año se pasa del número de ocupados al de puestos de trabajo equivalente a tiempo completo. Por lo demás, esta serie coincide exactamente con la del INE durante su período de solapamiento excepto en 1971, donde una de las dos series seguramente presenta un error tipográfico (el INE da 12.509 miles de ocupados y M&L 12.590). Puesto que la serie del INE está tomada directamente de su página web, he conservado el valor que ofrece esta fuente. Por lo demás, la serie de referencia que utilizo es la del INE, extendida hacia atrás hasta 1954 con la serie de M&L por yuxtaposición.

**Gráfico A1.3: Peso de la “extra-regio” en el empleo total de España**



**Gráfico A1.4: Empleo total, ratio FBBVA/CNEe86**



Como en el caso del VAB, la serie así obtenida se corrige a la baja excluyendo una estimación del empleo correspondiente a la extra-regio. El Gráfico A1.3 muestra la evolución del peso de este territorio ficticio en el empleo total de España entre 1980 y 1996. El valor de esta magnitud



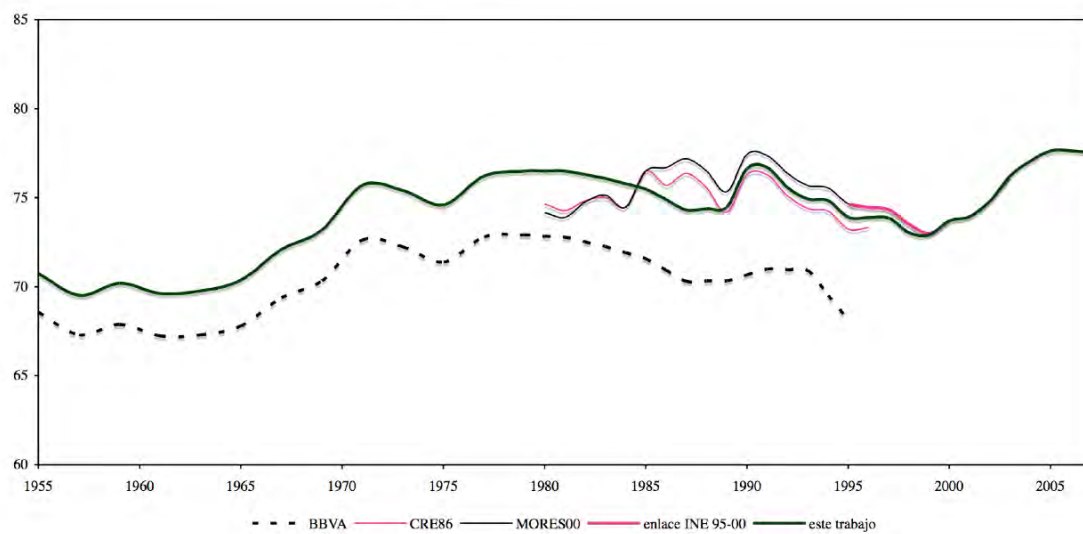
observado en 1980 (el 0.038%) se utiliza para corregir a la baja la serie original durante los años anteriores.

El Gráfico A1.4 compara la serie de empleo ajustada basada en la CNE con la serie análoga de la Fundación BBVA. Una vez más, la segunda se sitúa generalmente por encima de la primera, con diferencias que exceden el 10% en algunos años. En este caso, las mayores diferencias entre ambas series no se producen en los primeros años del período muestral sino durante la primera mitad de los ochenta.

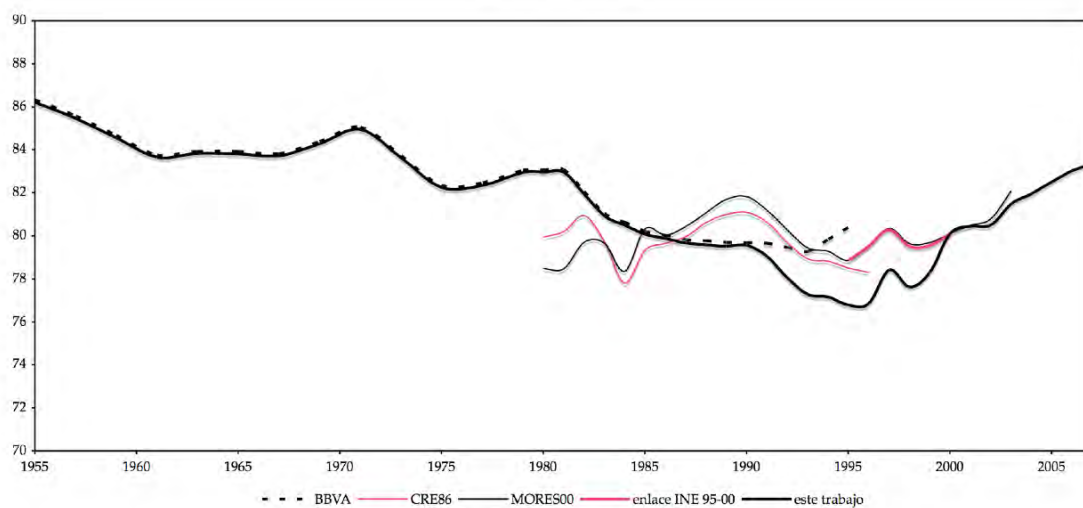
## Anexo 2: Gráficos comparativos

### Gráfico A2.1: Andalucía

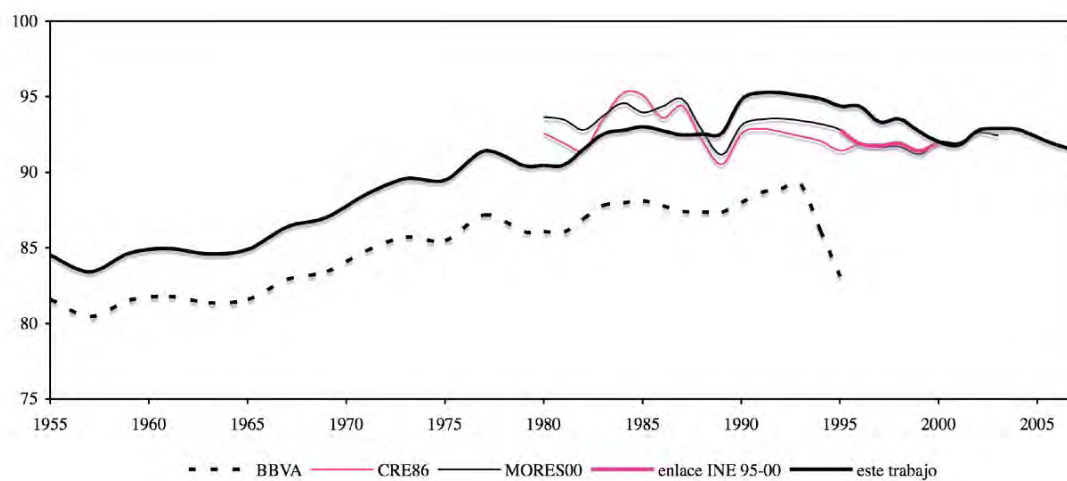
#### VAB per cápita a precios corrientes



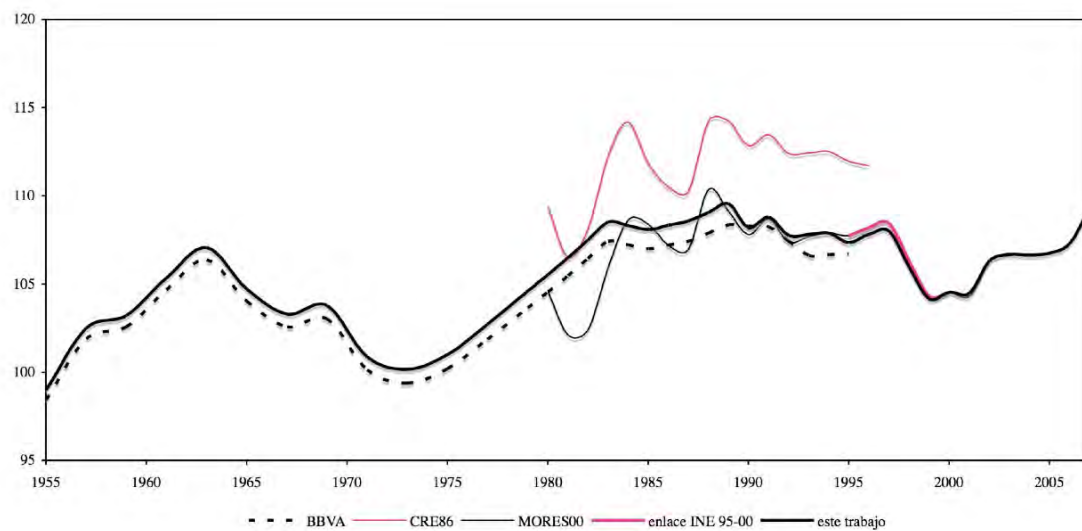
#### Empleos per cápita



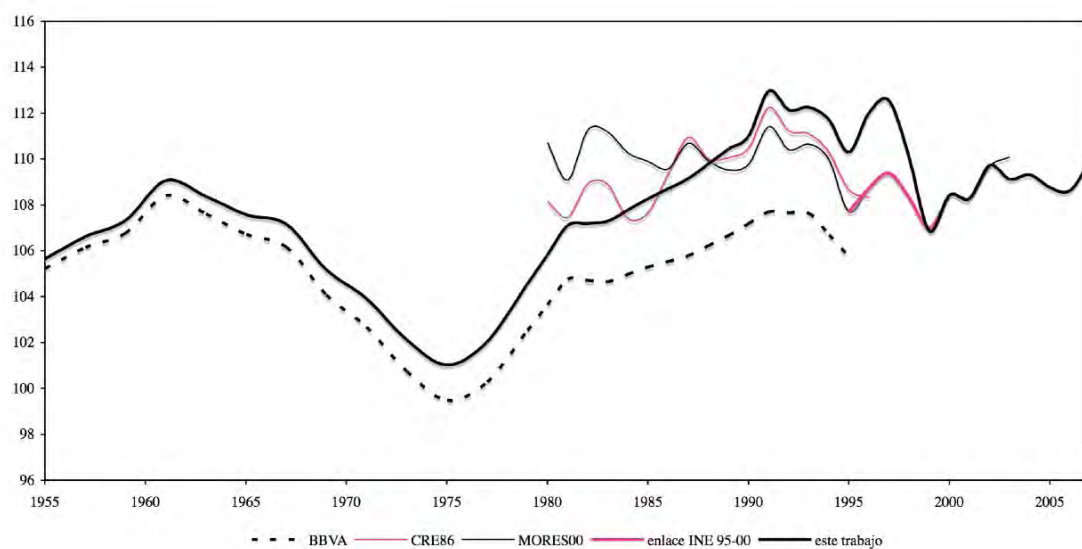
#### VAB por empleo a precios de 2000



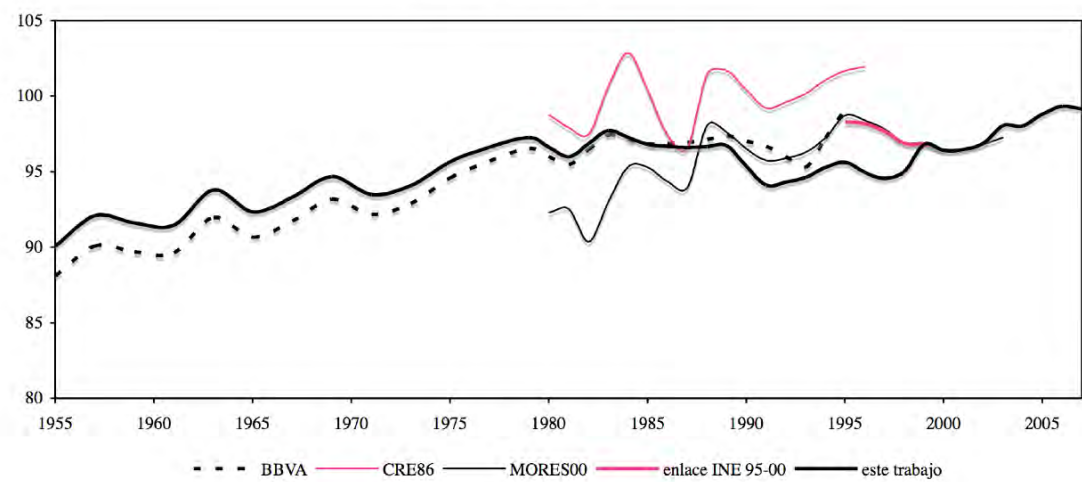
**Gráfico A2.2: Aragón**  
**VAB per cápita a precios corrientes**



**Empleos per cápita**

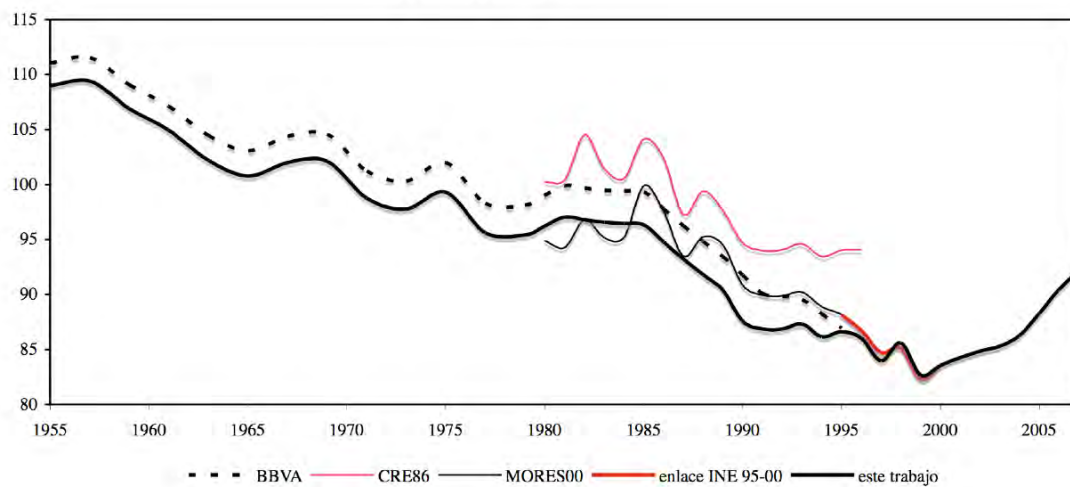


**VAB por empleo a precios de 2000**

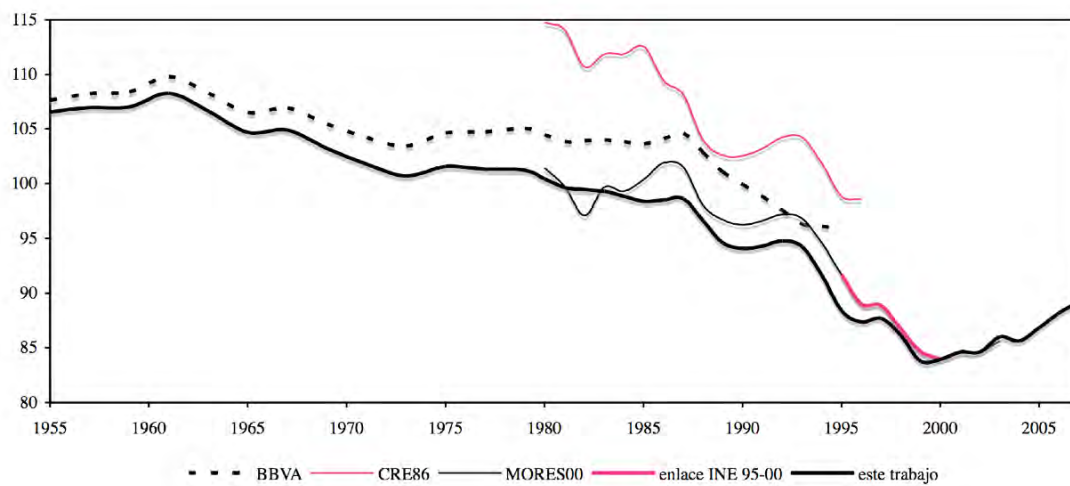


### Gráfico A2.3: Asturias

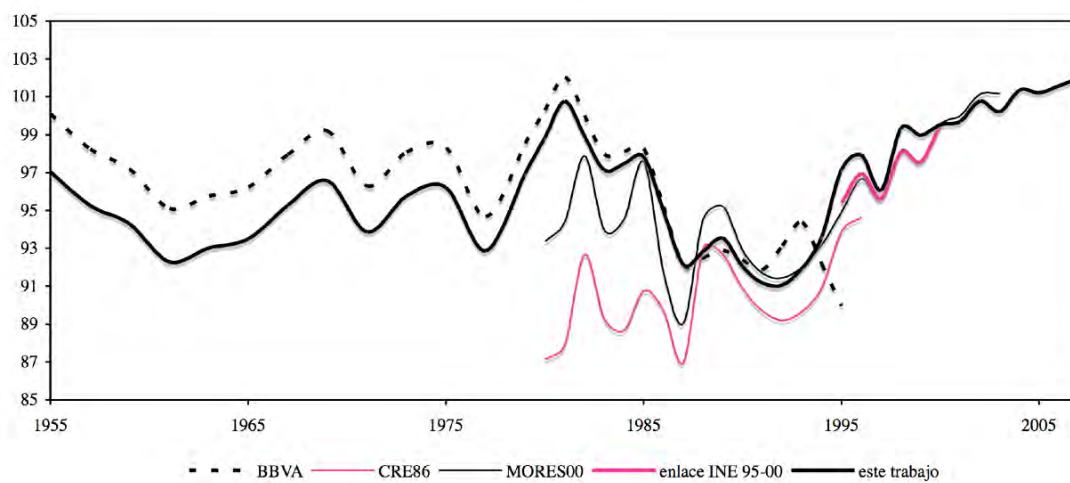
#### VAB per cápita a precios corrientes



#### Empleos per cápita

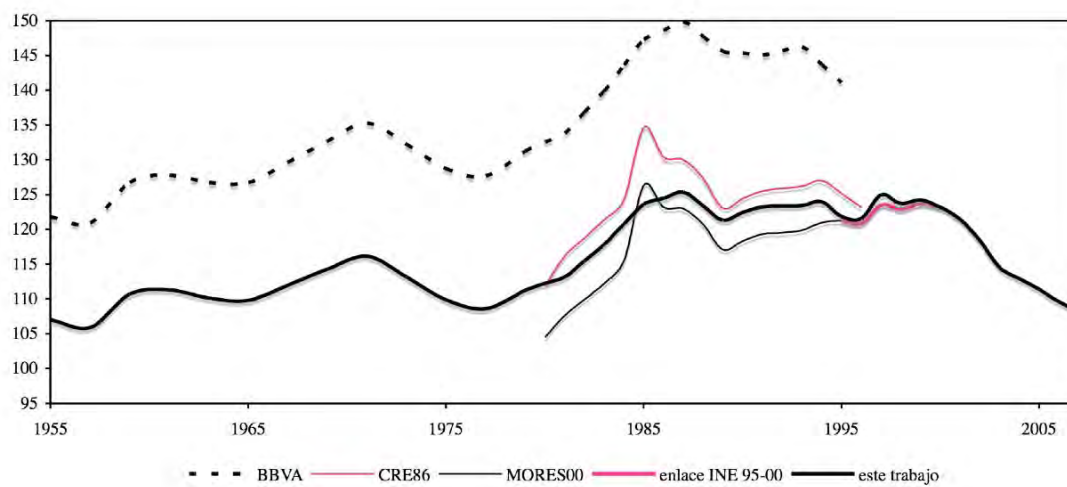


#### VAB por empleo a precios de 2000

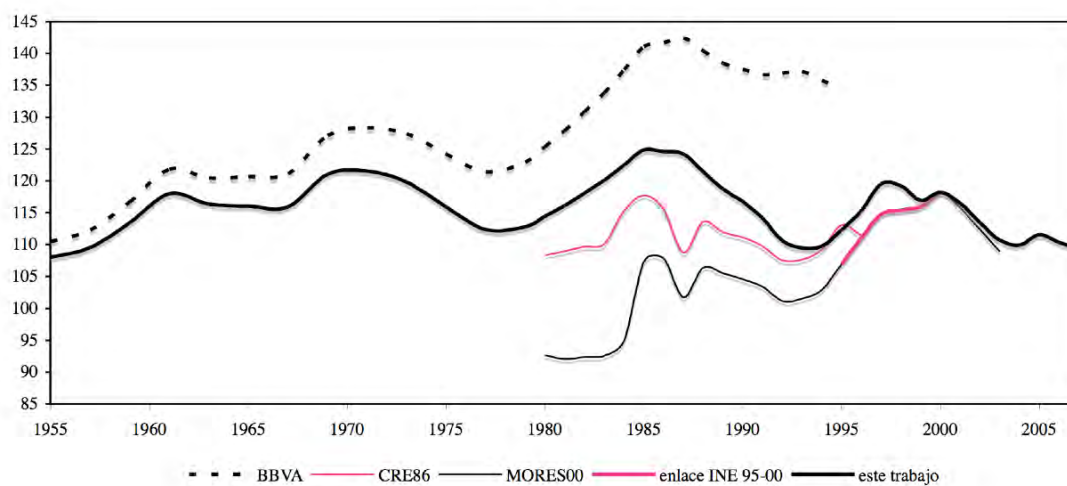


## Gráfico A2.4: Baleares

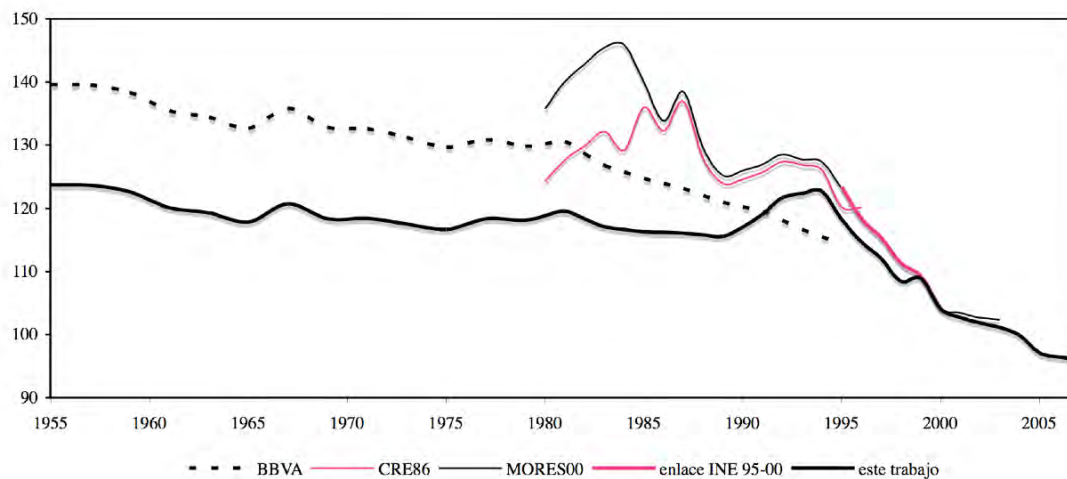
### VAB per cápita a precios corrientes



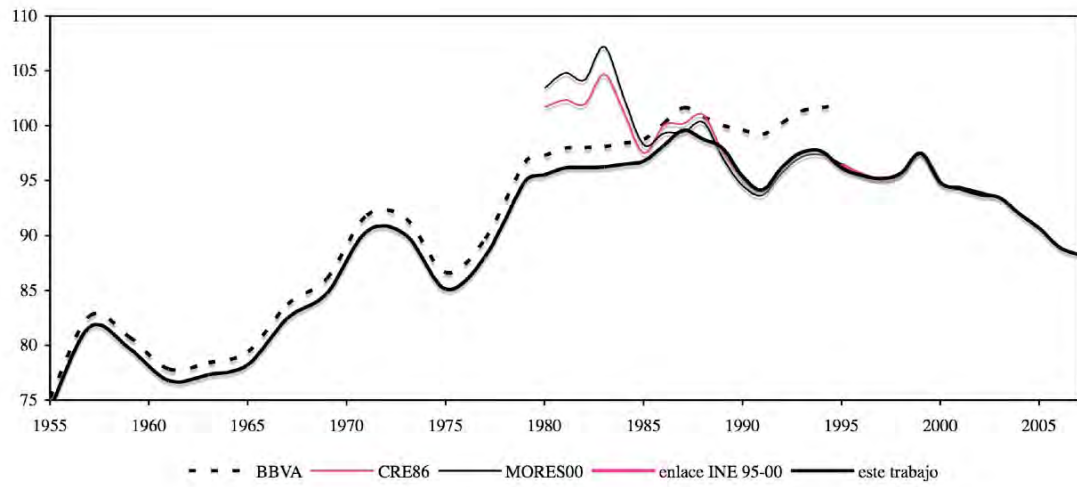
### Empleos per cápita



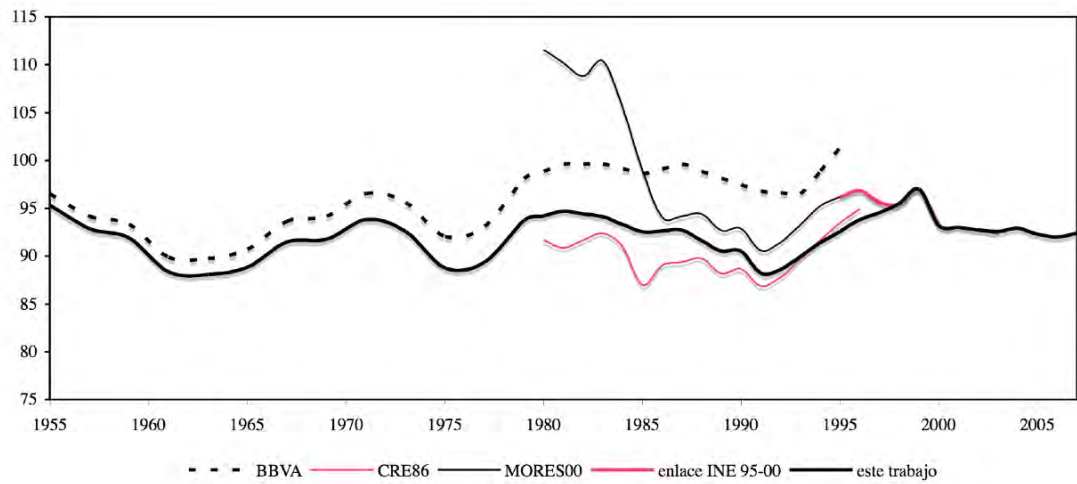
### VAB por empleo a precios de 2000



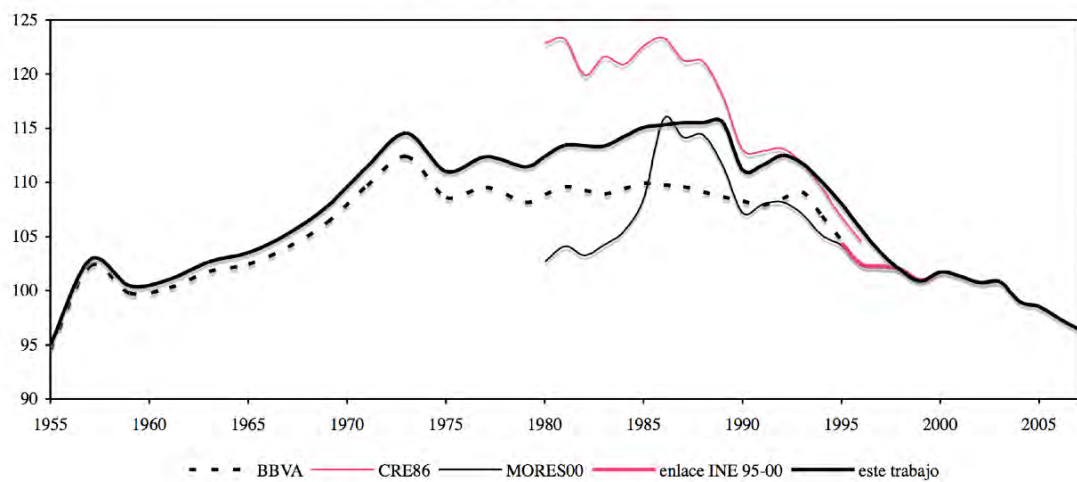
**Gráfico A2.5: Canarias**  
**VAB per cápita a precios corrientes**



**Empleos per cápita**



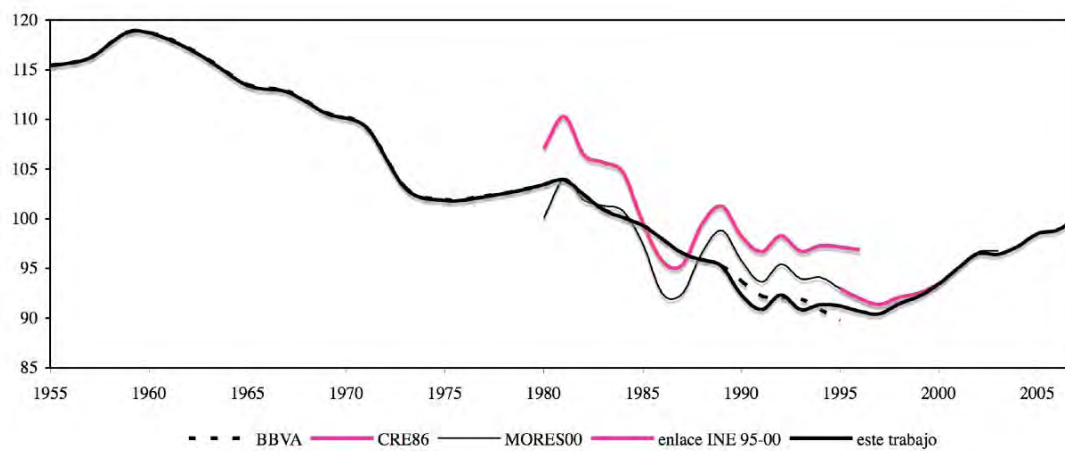
**VAB por empleo a precios de 2000**



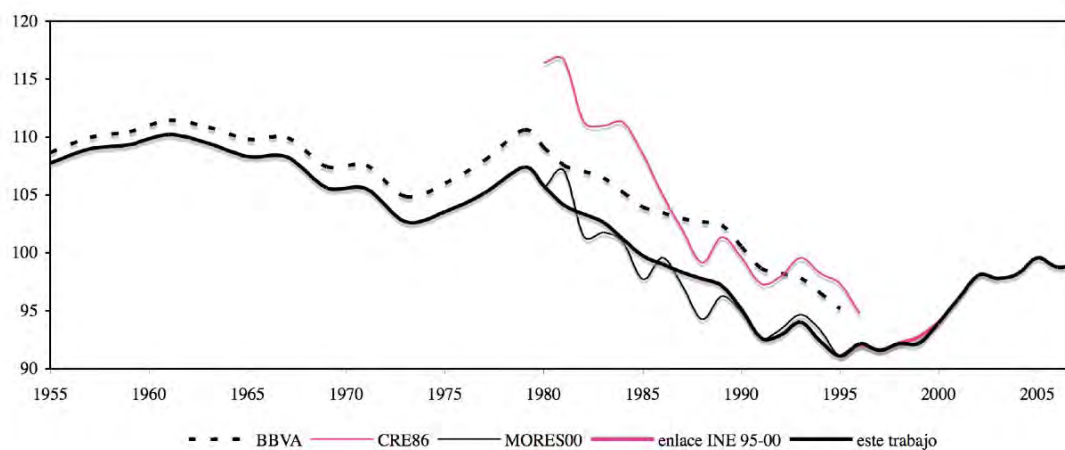


## Gráfico A2.6: Cantabria

### VAB per cápita a precios corrientes



### Empleos per cápita



### VAB por empleo a precios de 2000

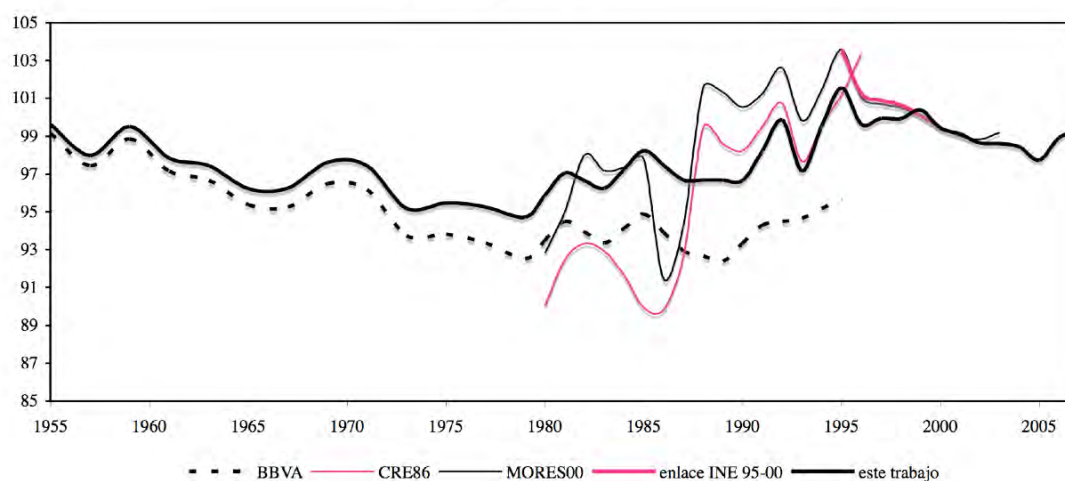
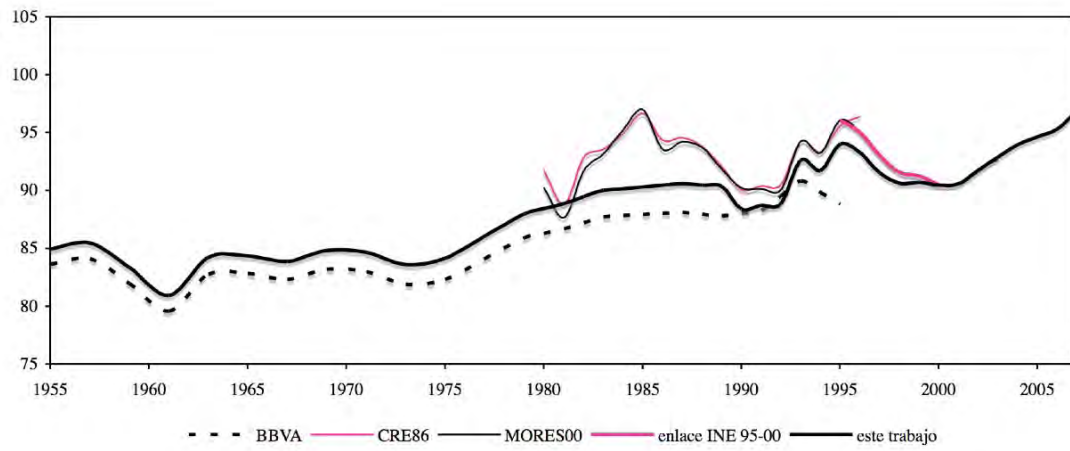
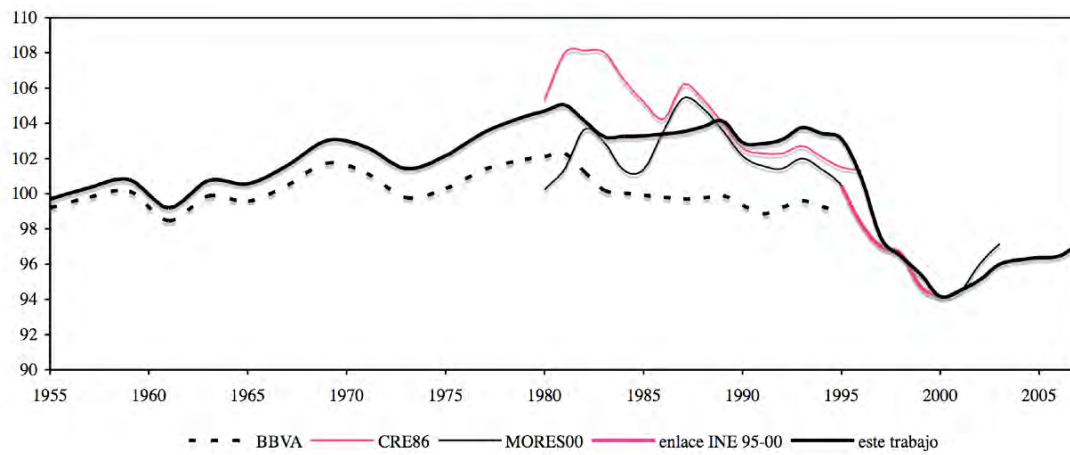


Gráfico A2.7: Castilla y León  
VAB per cápita a precios corrientes



Empleos per cápita



VAB por empleo a precios de 2000

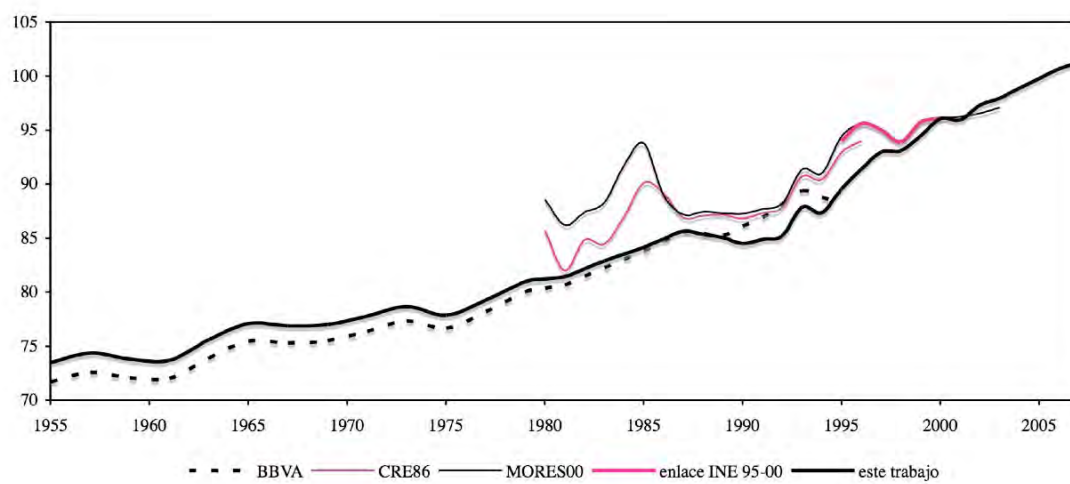
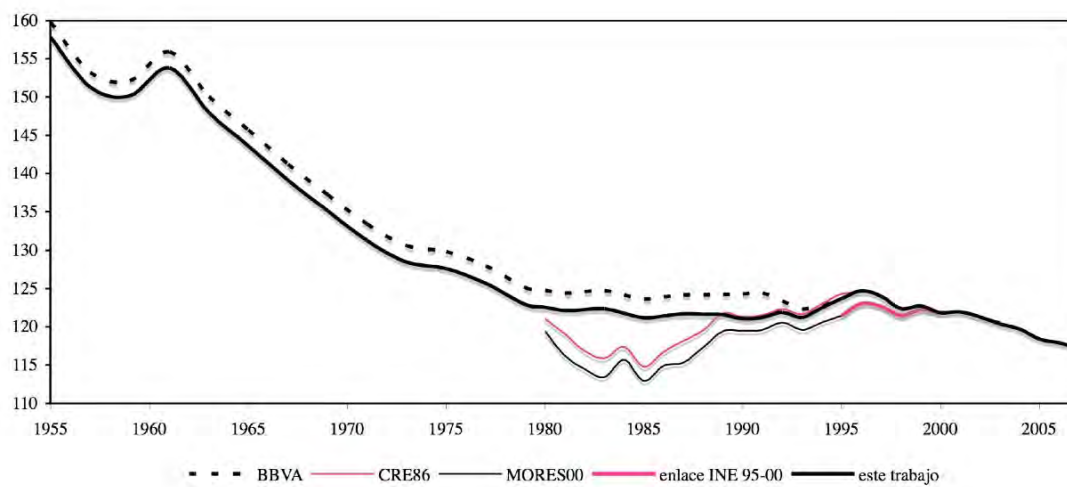
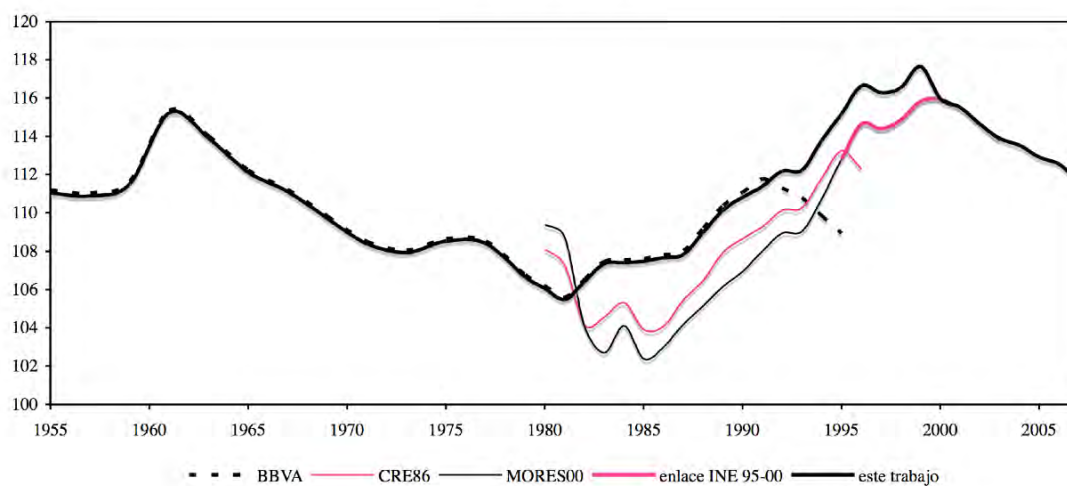




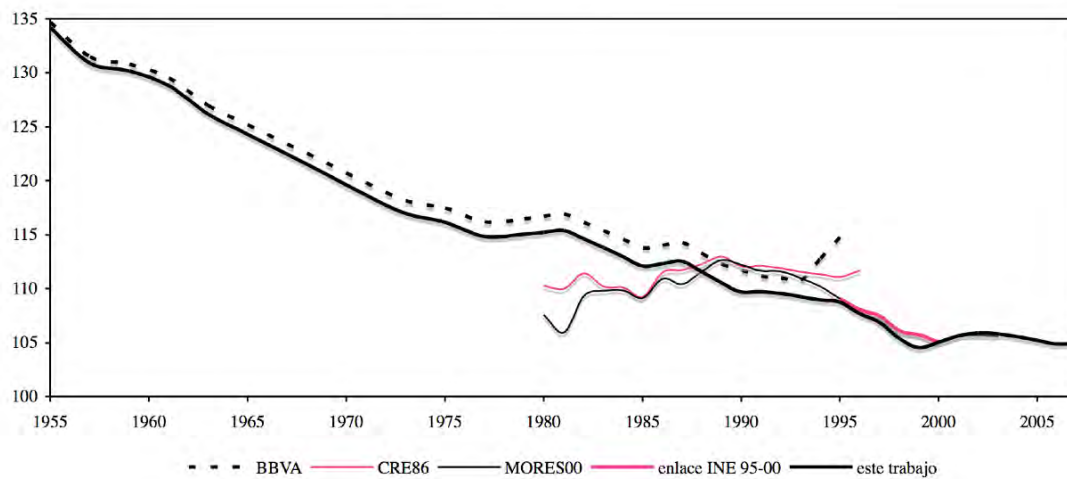
Gráfico A2.9: Cataluña  
VAB per cápita a precios corrientes



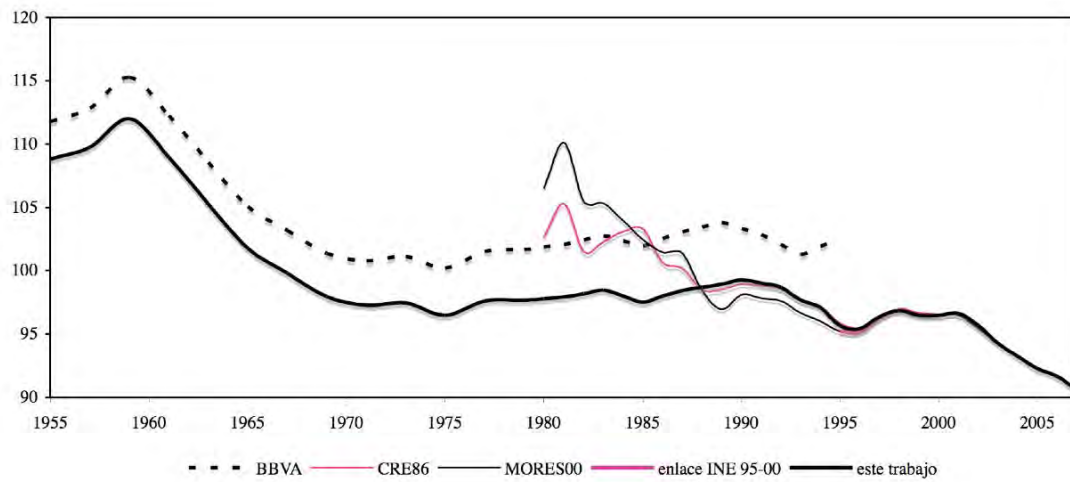
Empleos per cápita



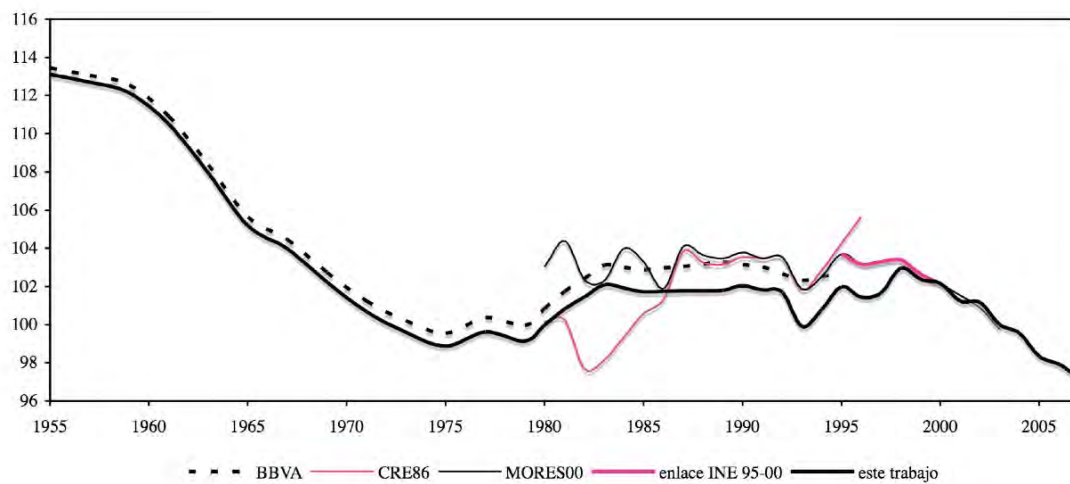
VAB por empleo a precios de 2000



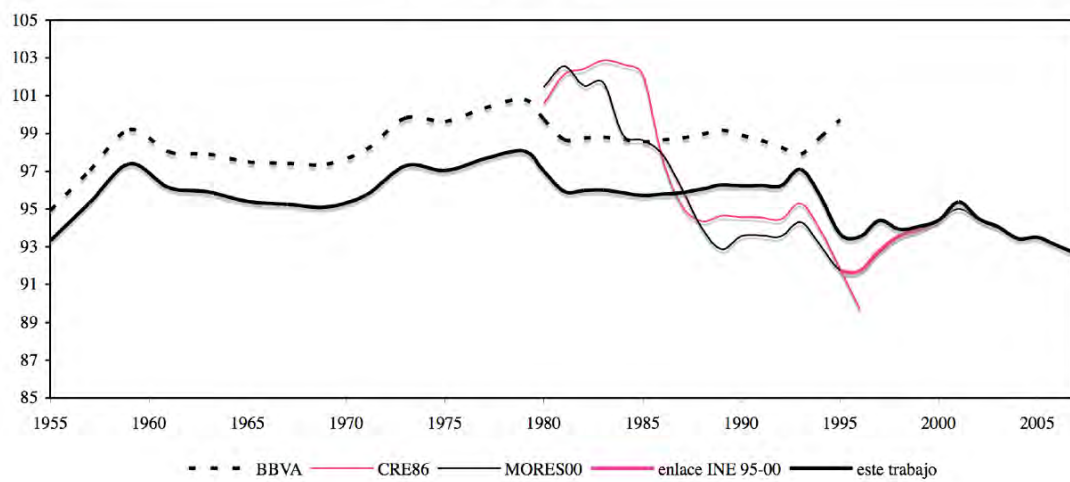
**Gráfico A2.10: Valencia**  
**VAB per cápita a precios corrientes**



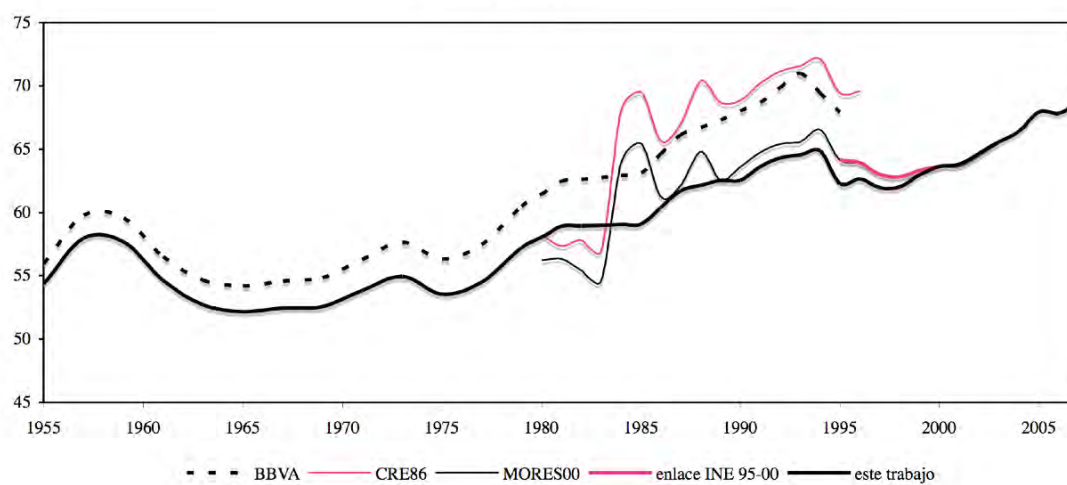
**Empleos per cápita**



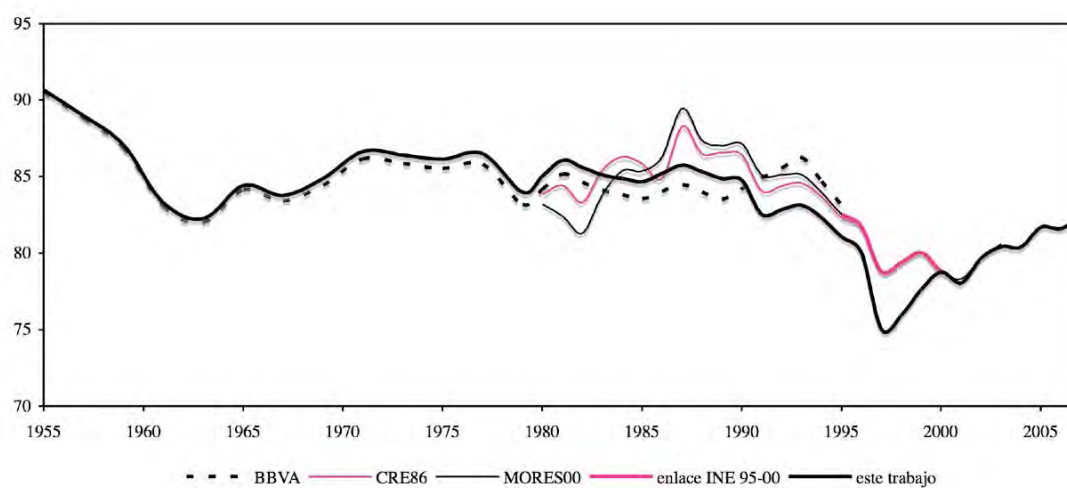
**VAB por empleo a precios de 2000**



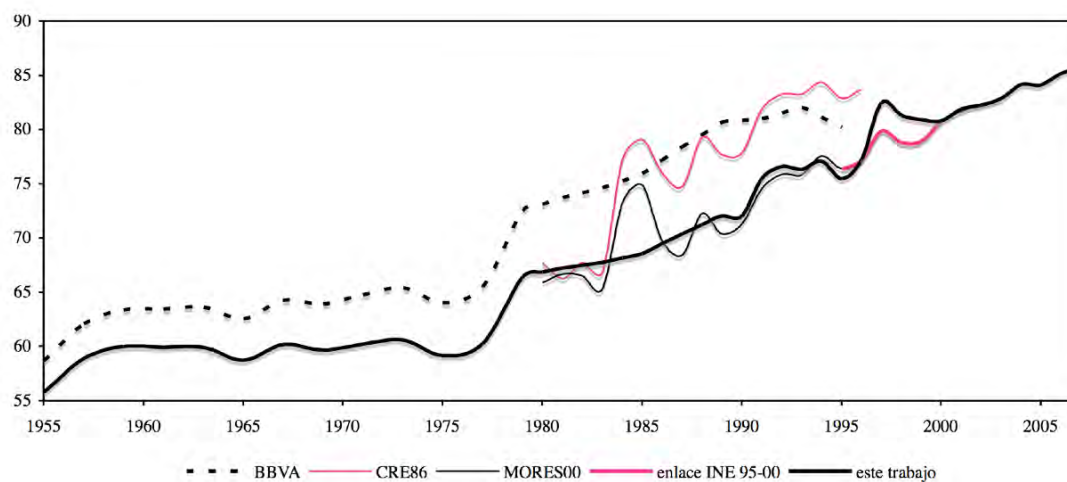
**Gráfico A2.11: Extremadura**  
**VAB per cápita a precios corrientes**



**Empleos per cápita**

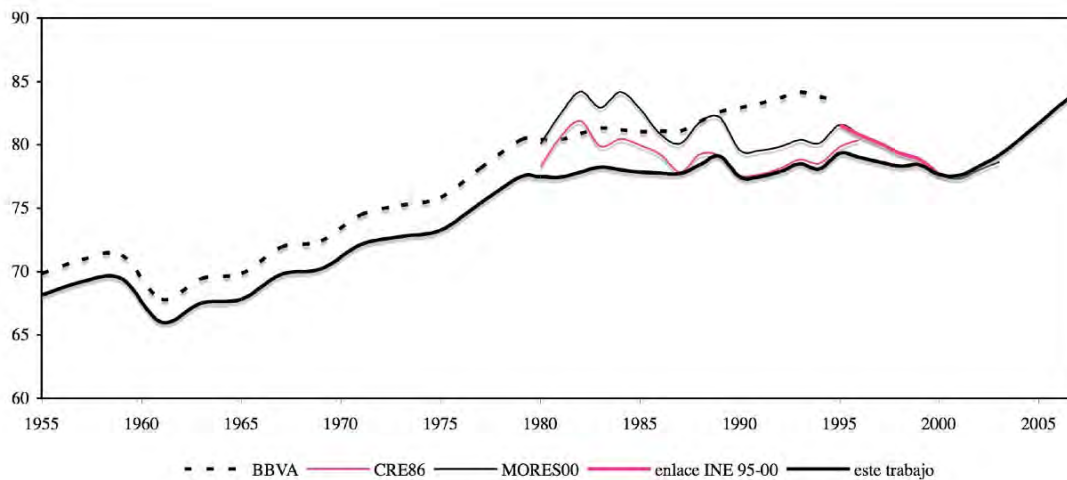


**VAB por empleo a precios de 2000**

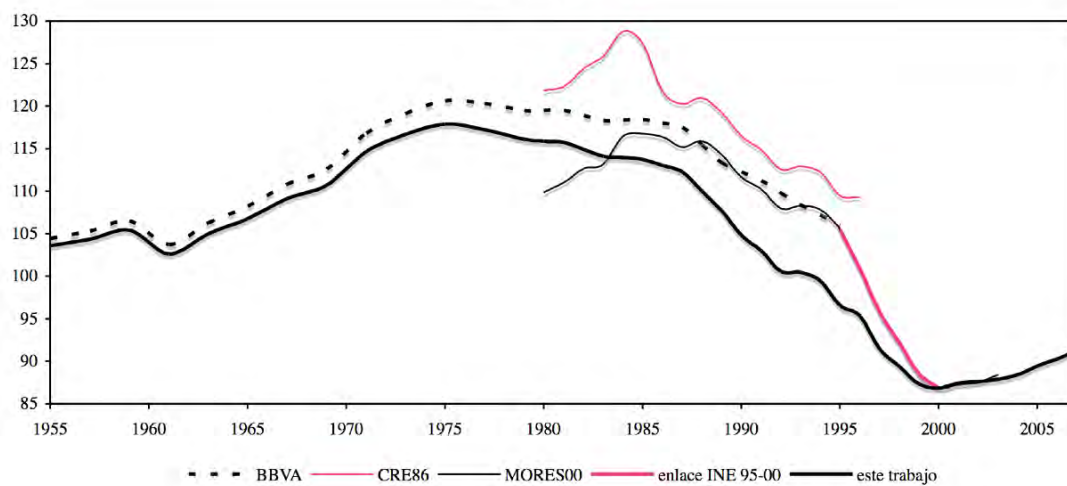


### Gráfico A2.12: Galicia

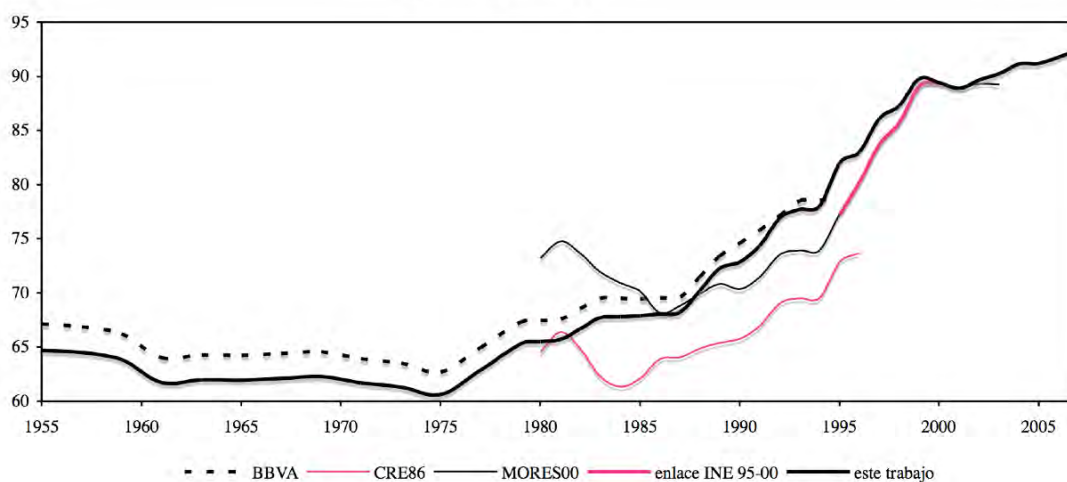
#### VAB per cápita a precios corrientes



#### Empleos per cápita

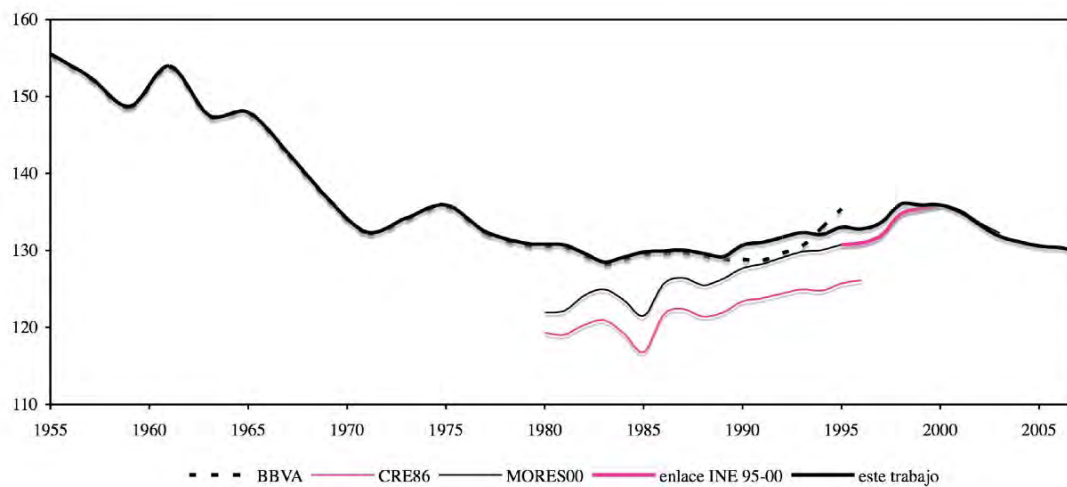


#### VAB por empleo a precios de 2000

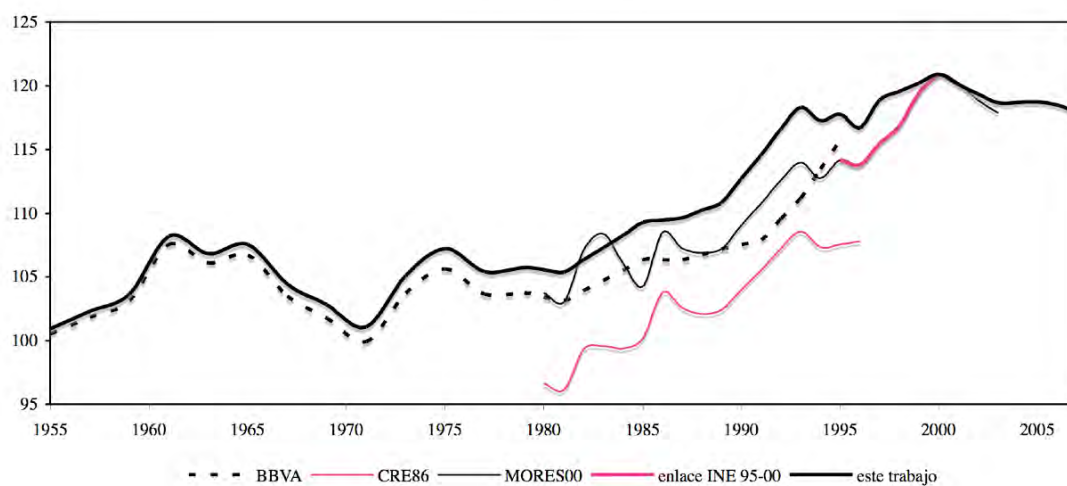


### Gráfico A2.13: Madrid

#### VAB per cápita a precios corrientes



#### Empleos per cápita



#### VAB por empleo a precios de 2000

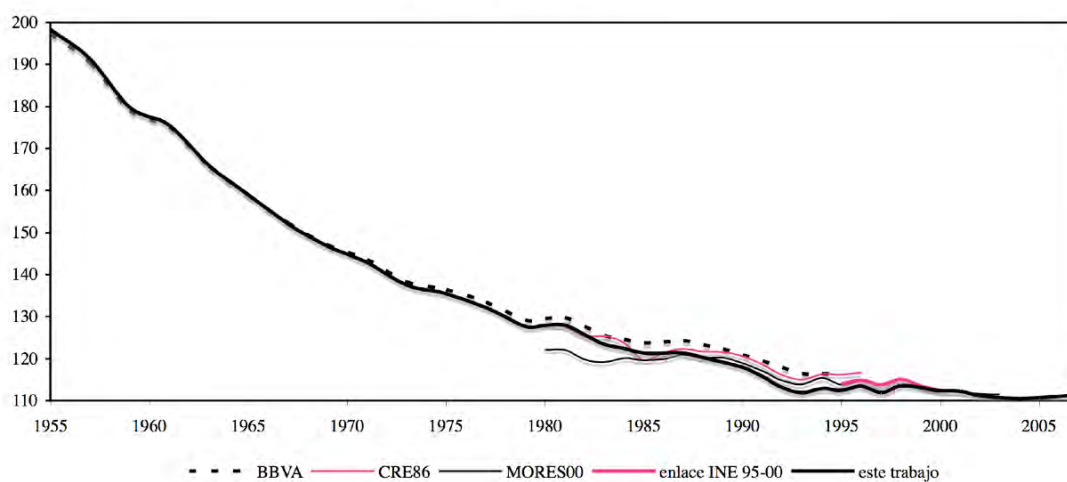
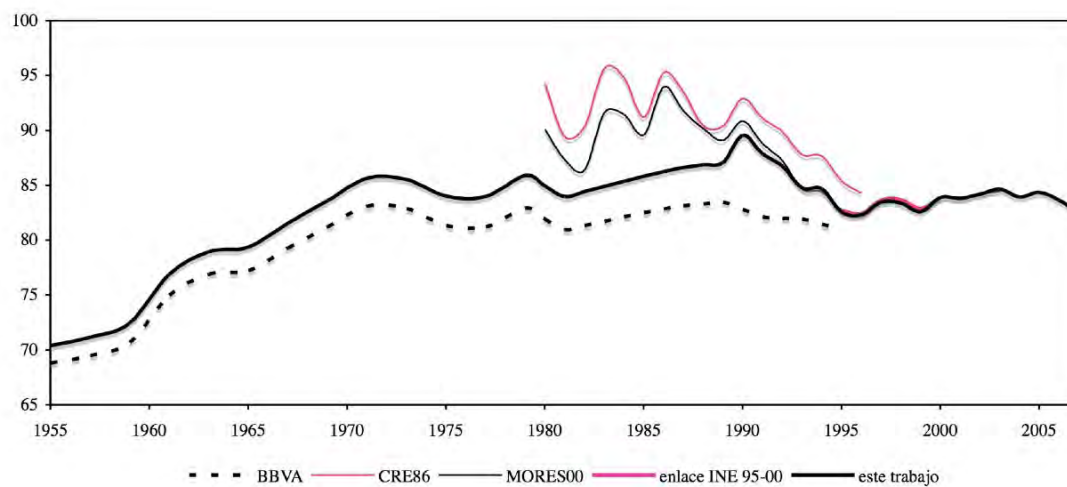
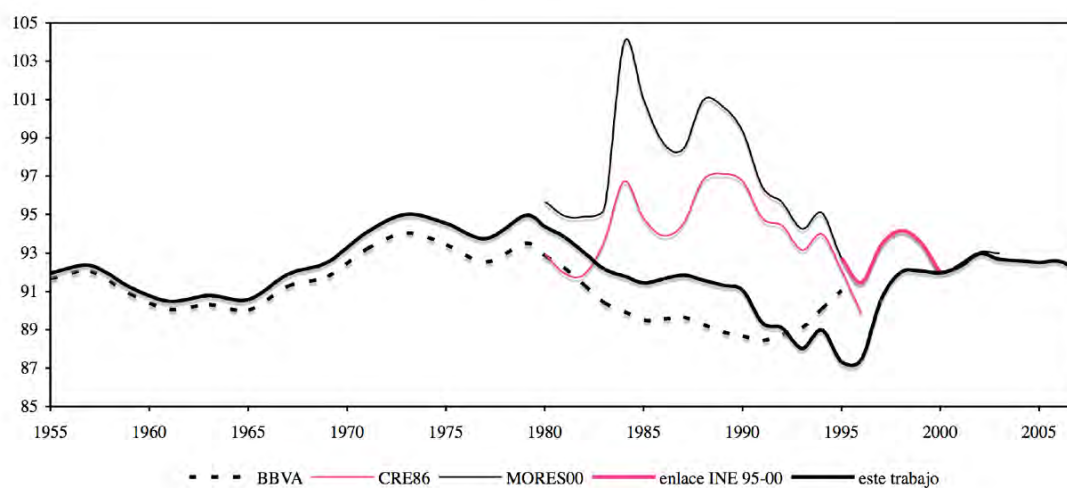


Gráfico A2.14: Murcia

VAB per cápita a precios corrientes



Empleos per cápita



VAB por empleo a precios de 2000

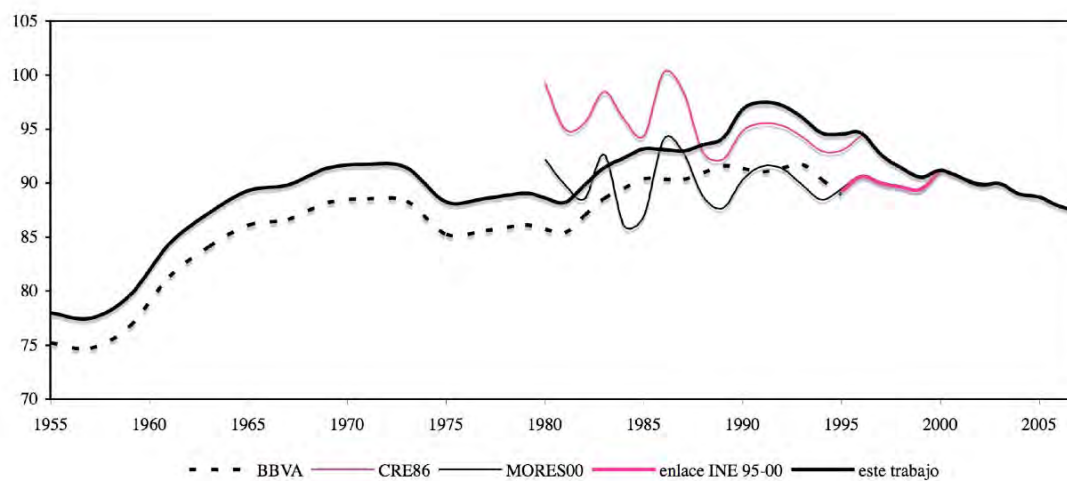
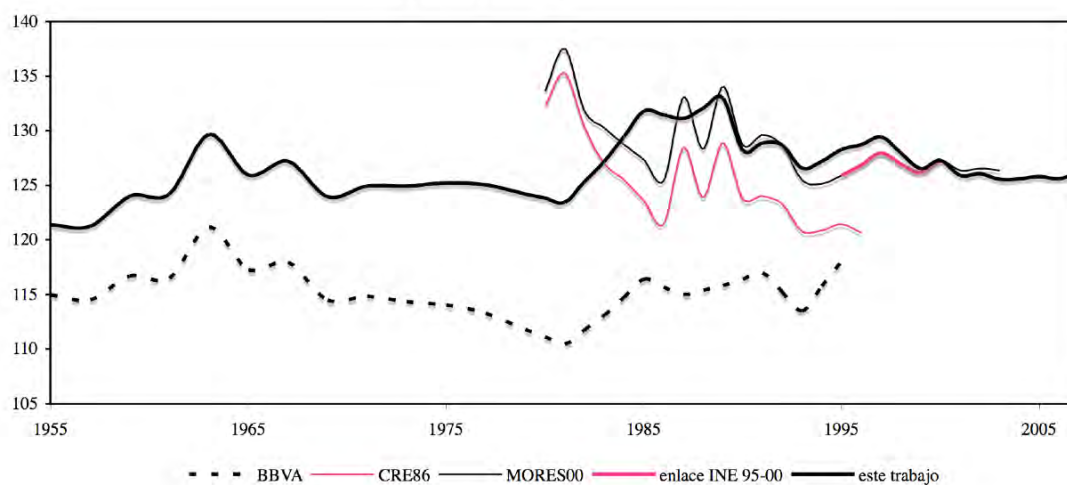


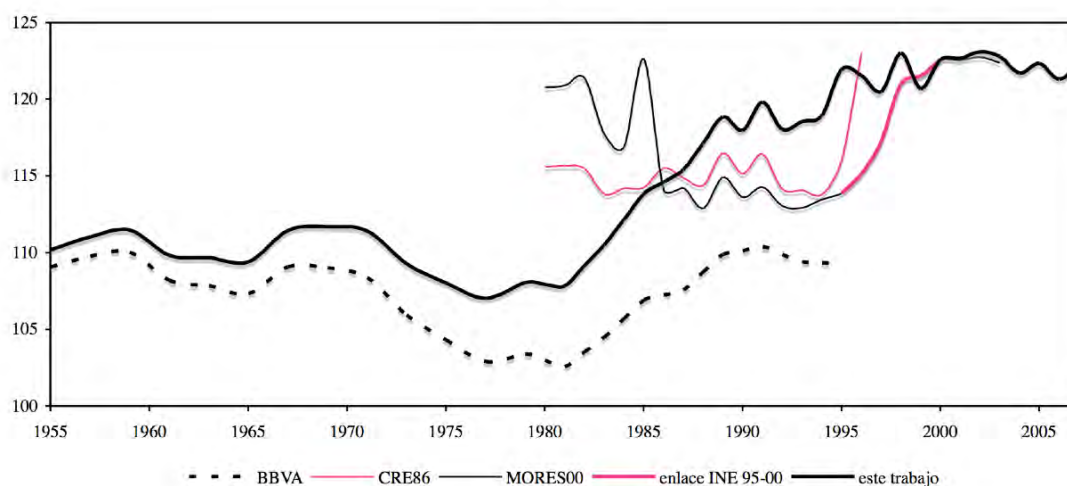


Gráfico A2.15: Navarra

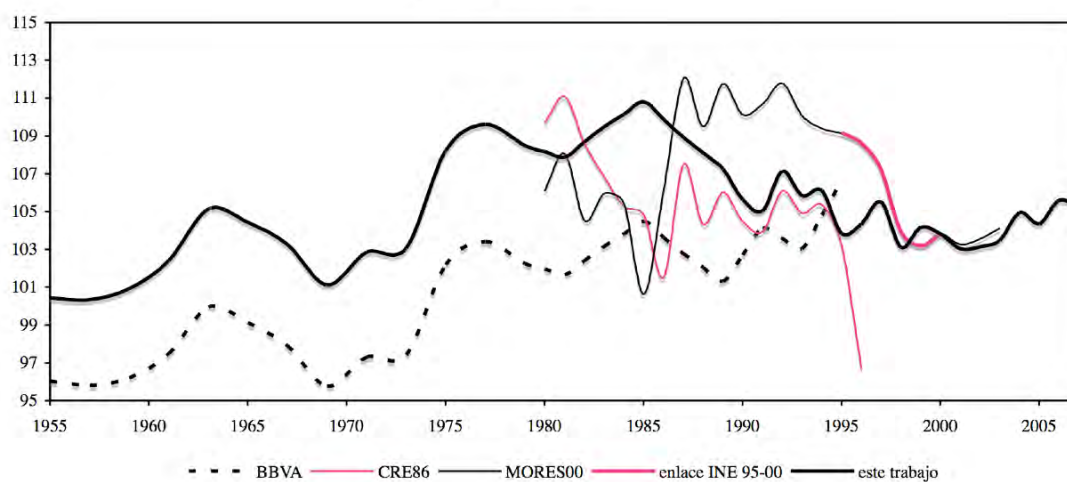
VAB per cápita a precios corrientes



Empleos per cápita

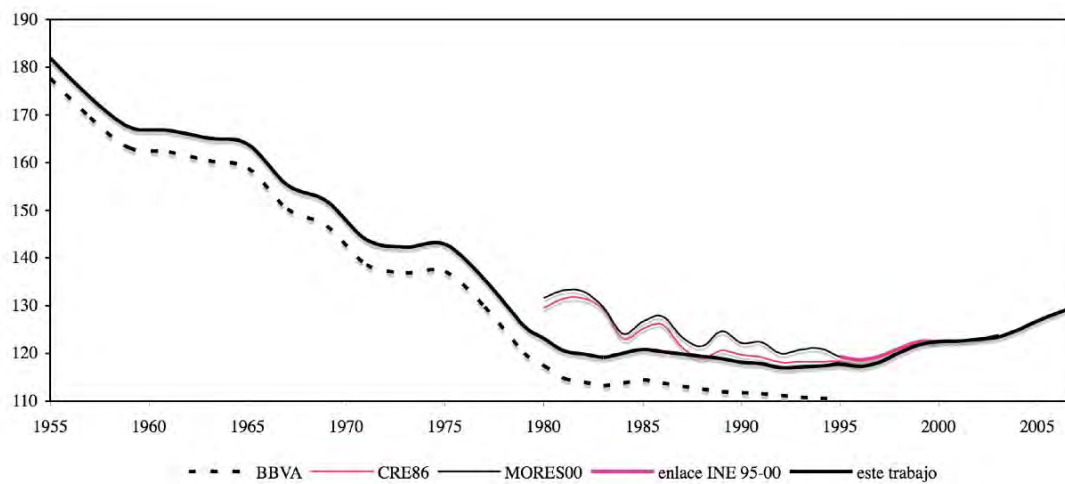


VAB por empleo a precios de 2000

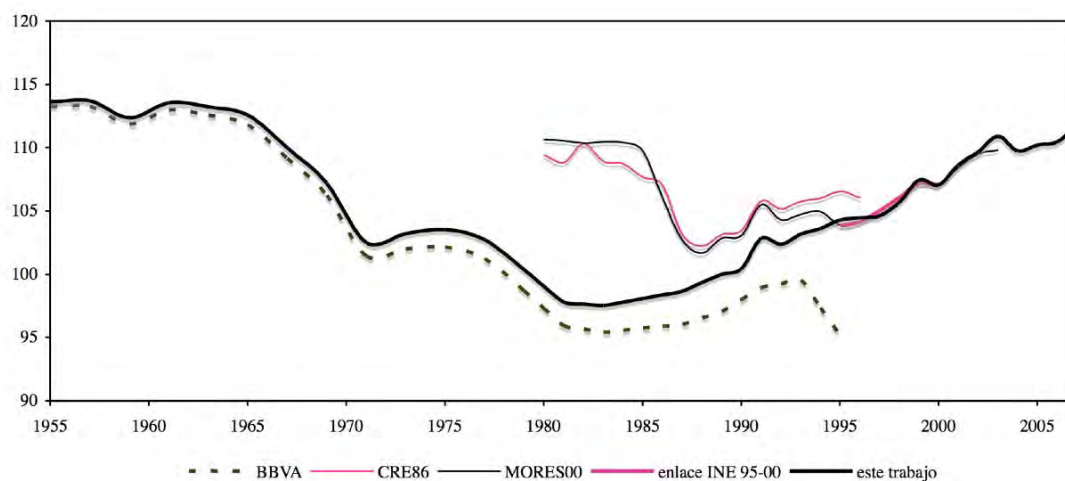


## Gráfico A2.16: País Vasco

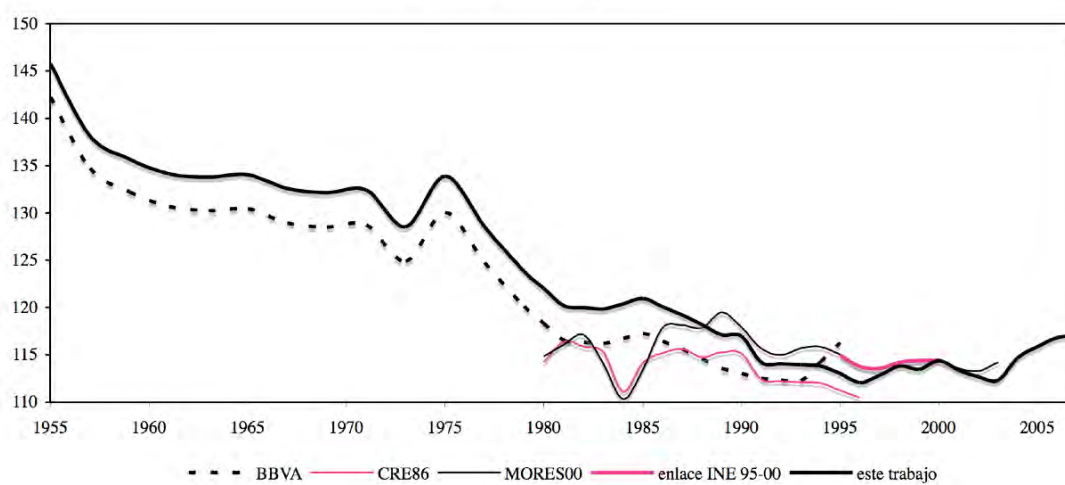
### VAB per cápita a precios corrientes



### Empleos per cápita



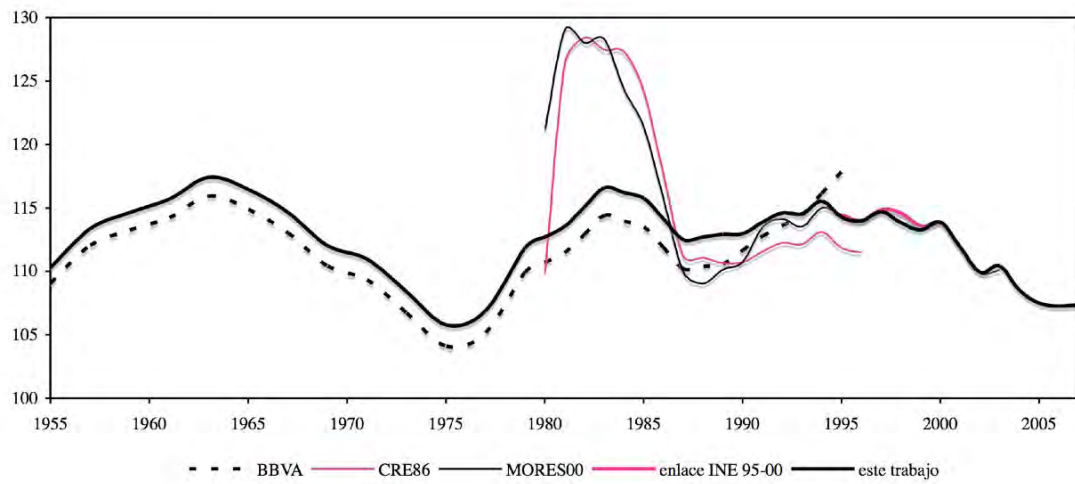
### VAB por empleo a precios de 2000



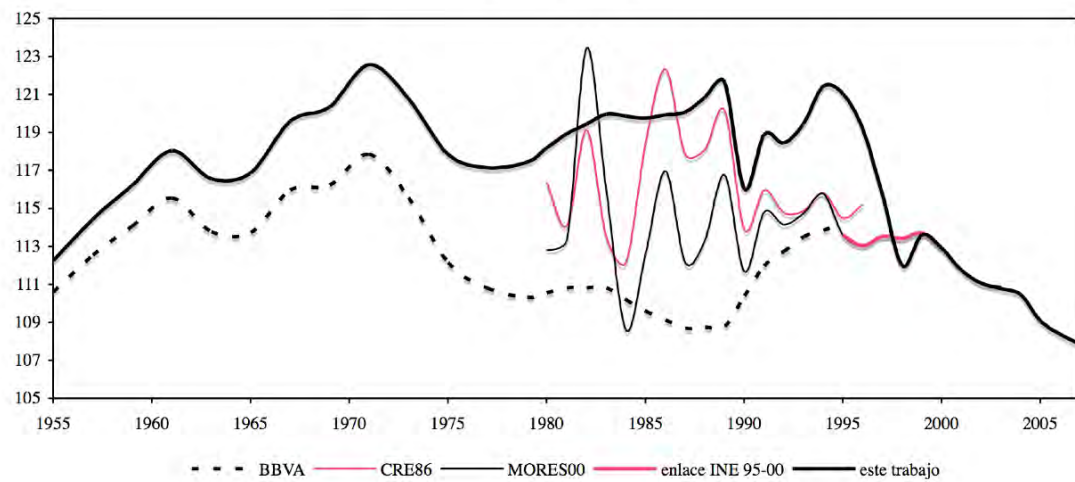


### Gráfico A2.17: Rioja

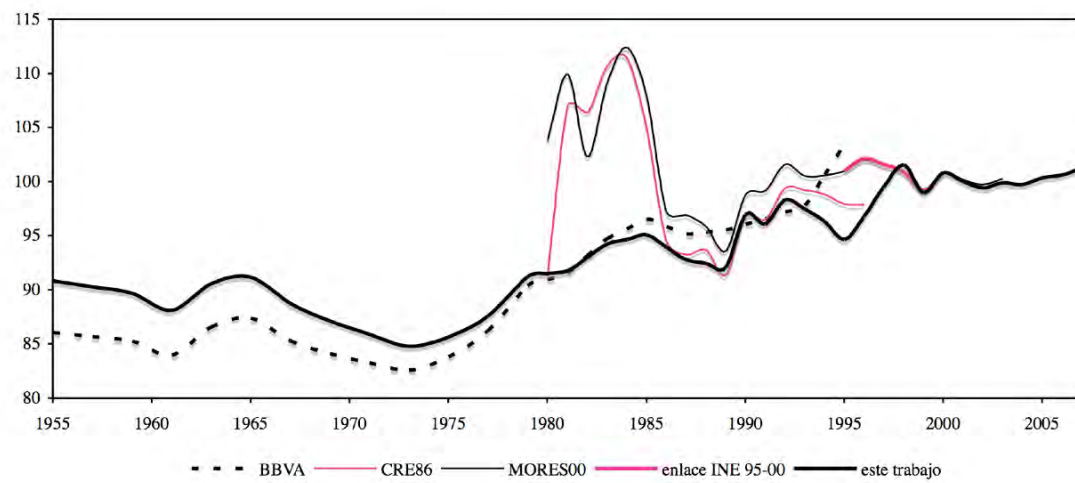
#### VAB per cápita a precios corrientes



#### Empleos per cápita

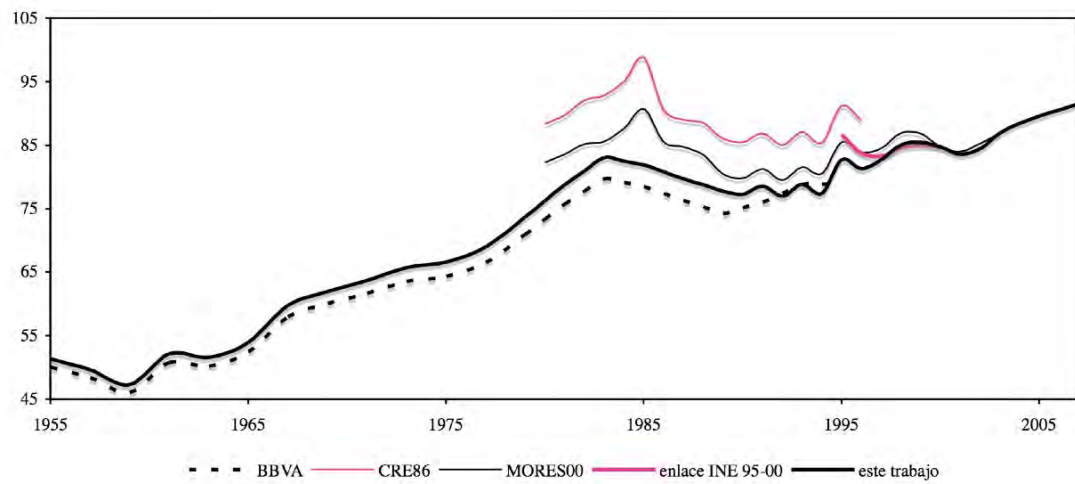


#### VAB por empleo a precios de 2000

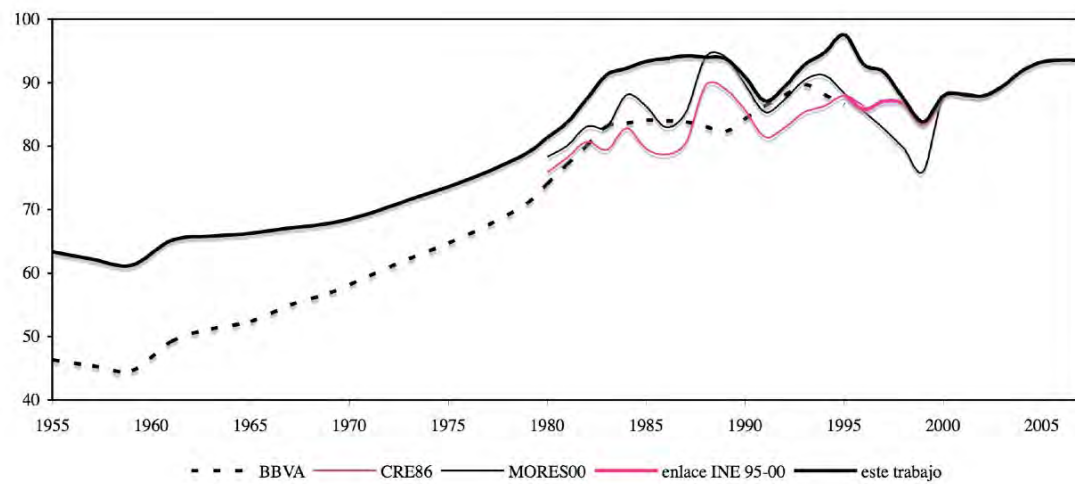


## Gráfico A2.18: Ceuta y Melilla

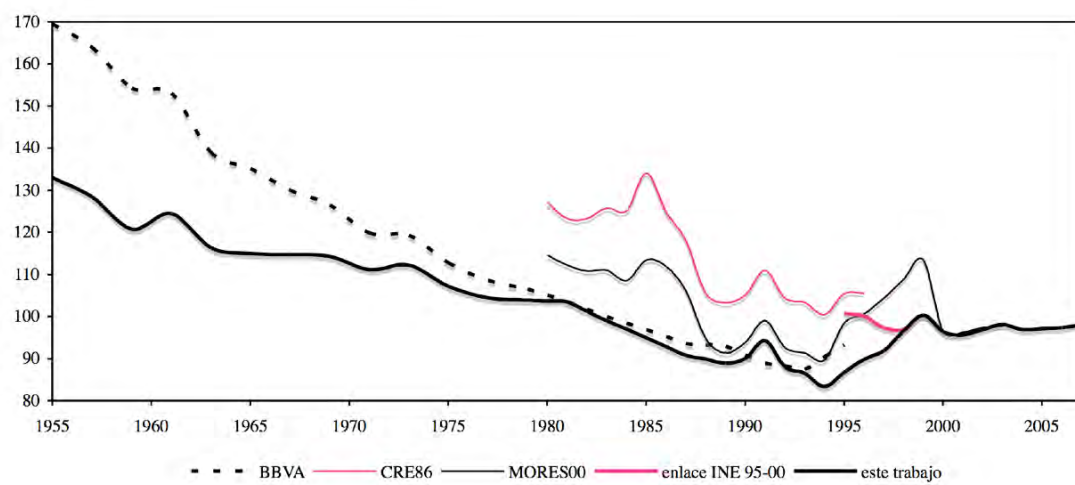
### VAB per cápita a precios corrientes



### Empleos per cápita

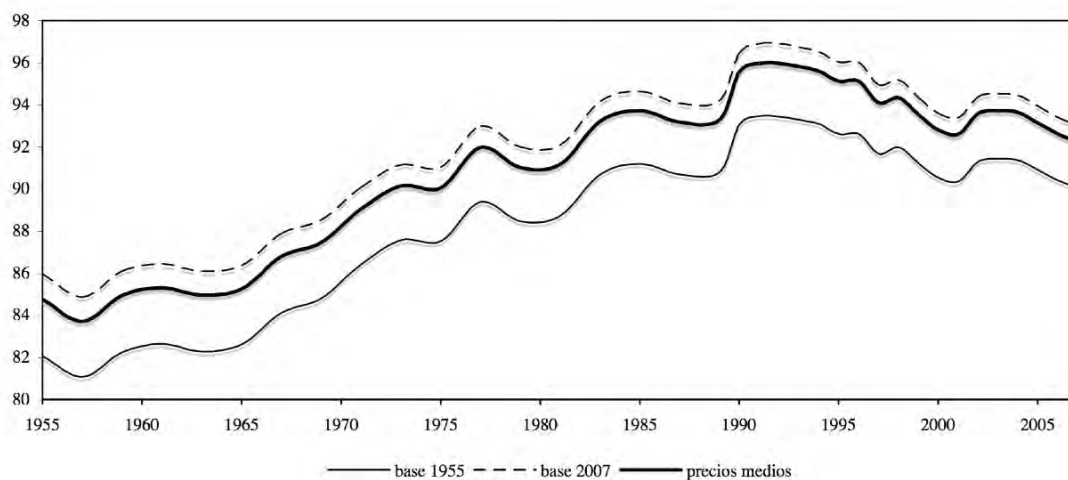


### VAB por empleo a precios de 2000

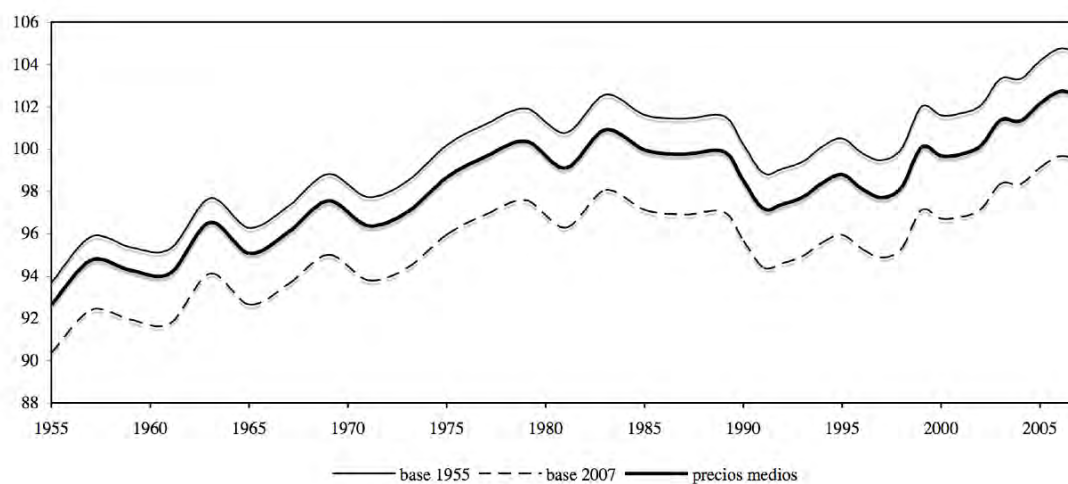


**Gráfico A2.19: Productividad (VAB por empleo a precios constantes)  
deflactado con distintos índices de precios**

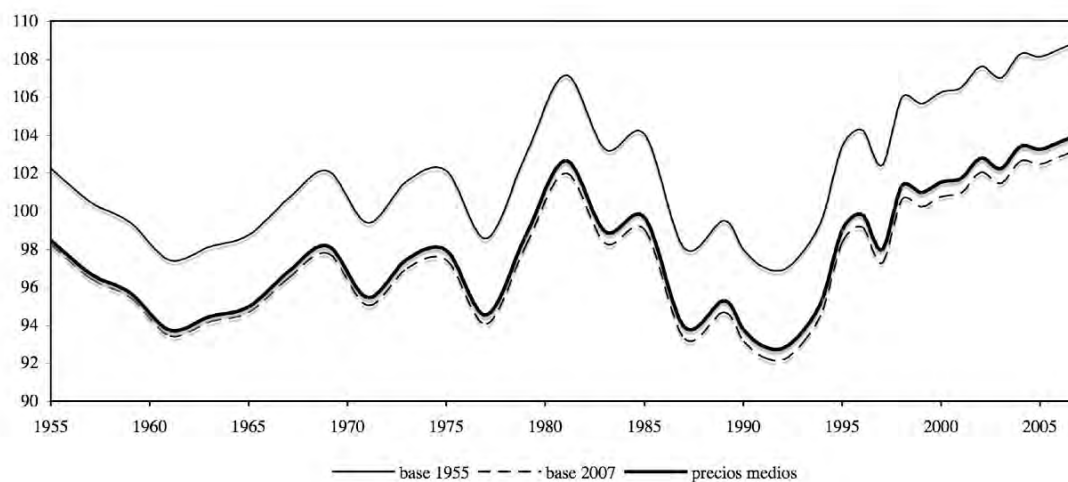
**Andalucía**



**Aragón**

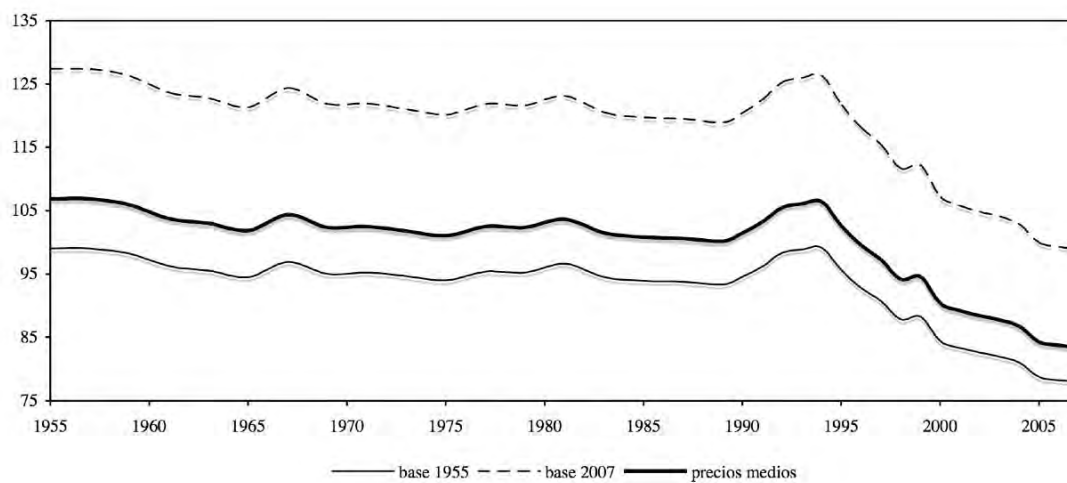


**Asturias**

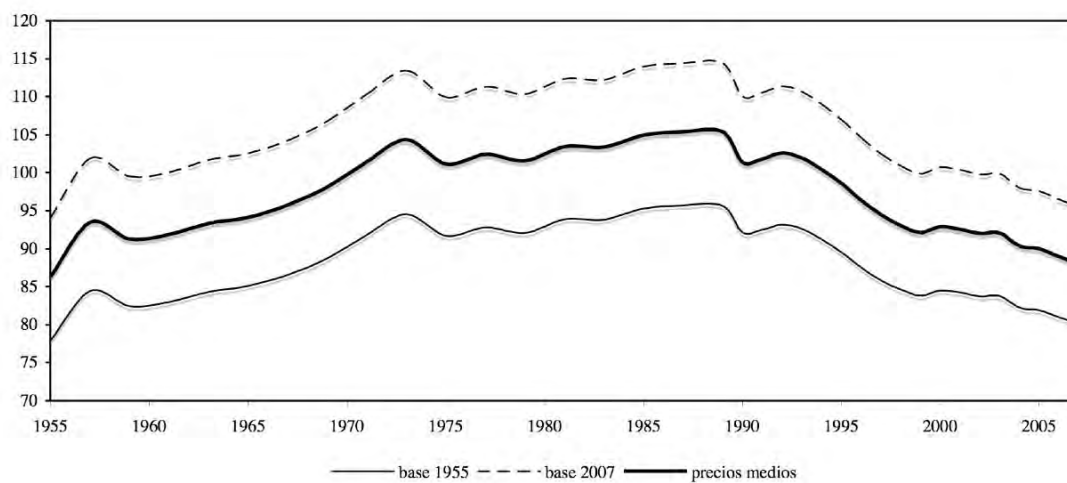


**Gráfico A2.19: Productividad (VAB por empleo a precios constantes)  
deflactado con distintos índices de precios -- continuación**

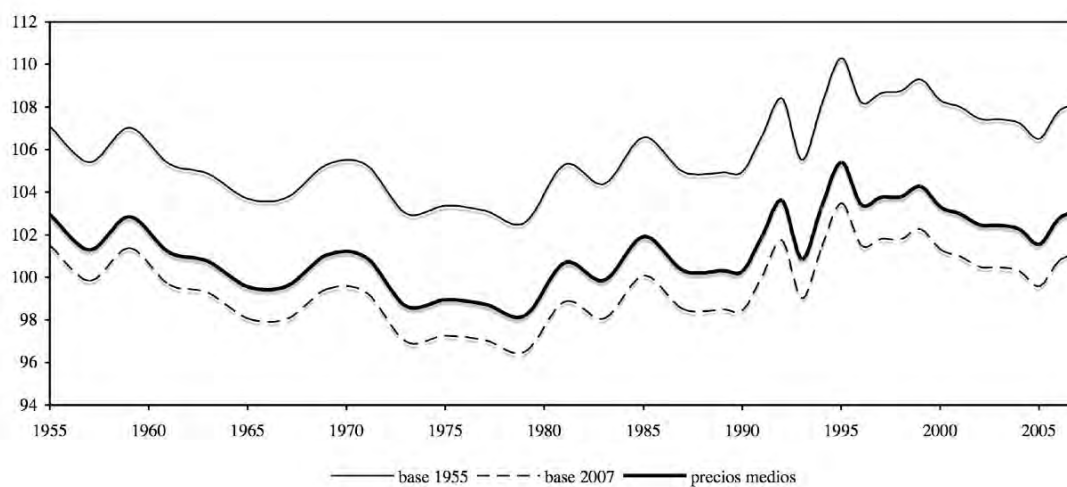
**Baleares**



**Canarias**

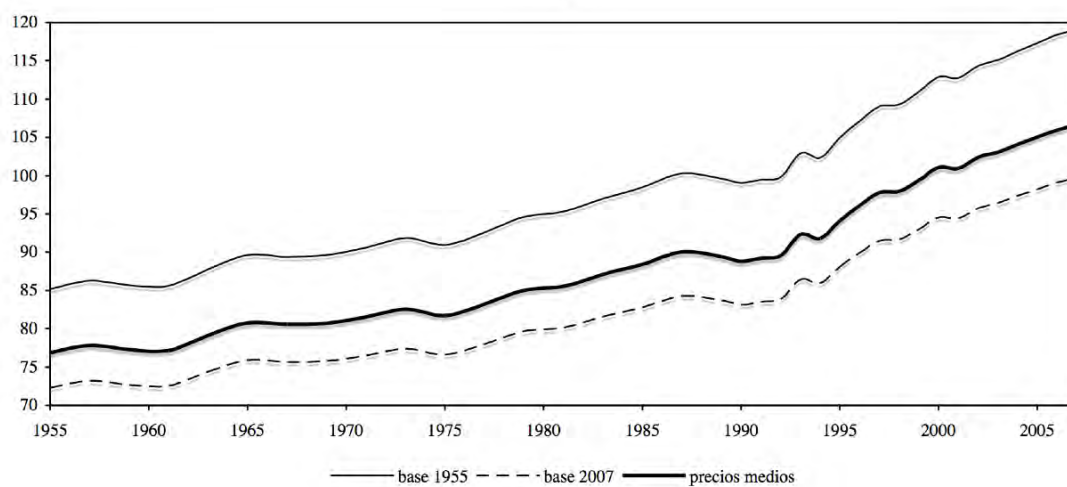


**Cantabria**

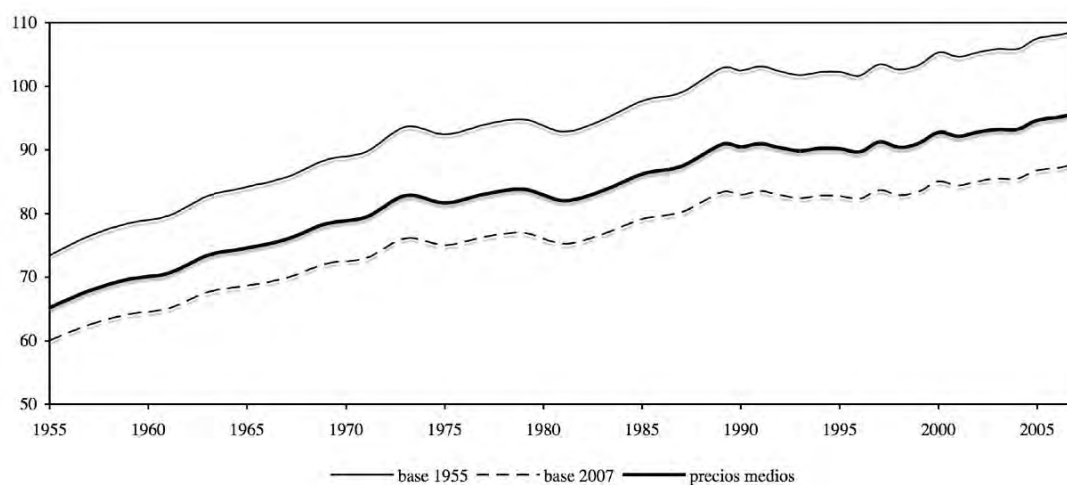


**Gráfico A2.19: Productividad (VAB por empleo a precios constantes)  
deflactado con distintos índices de precios -- continuación**

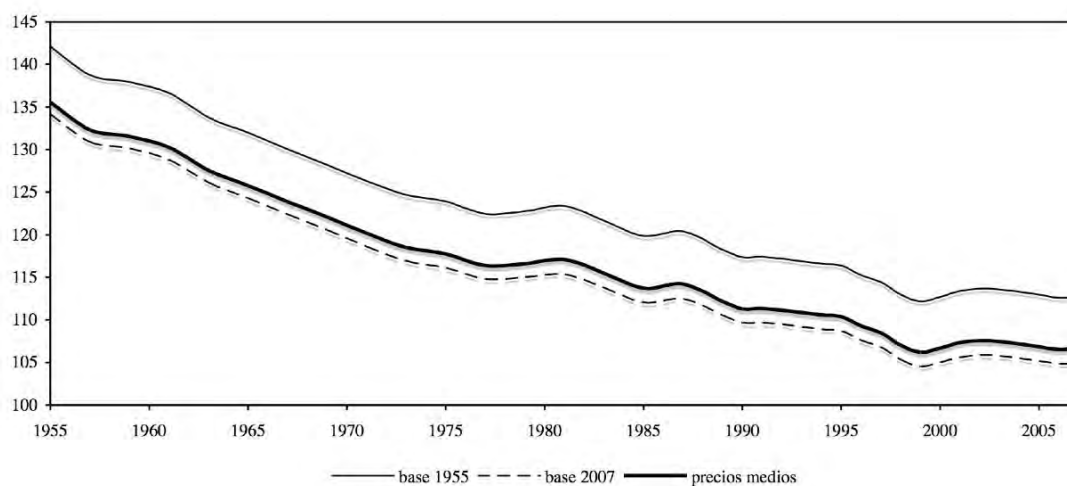
### Castilla y León



### Castilla la Mancha

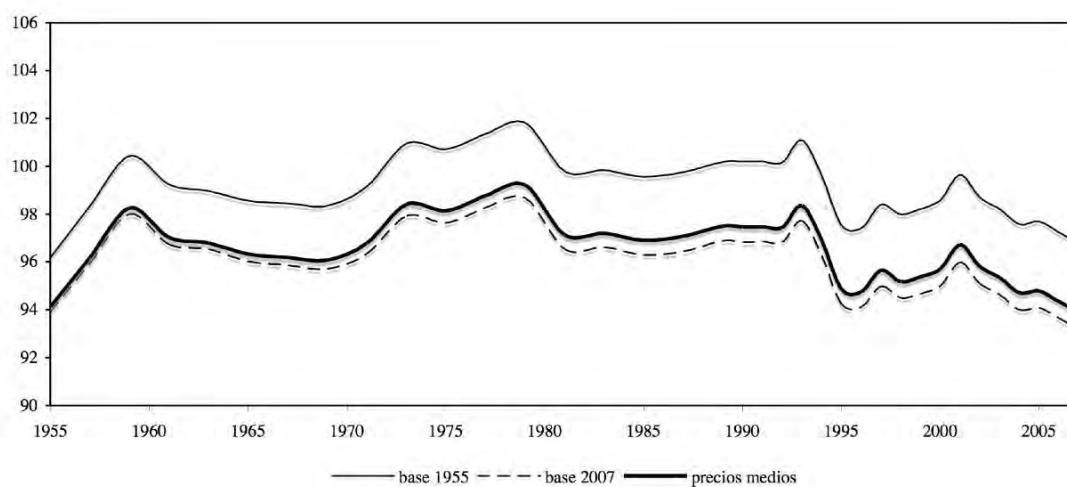


### Cataluña

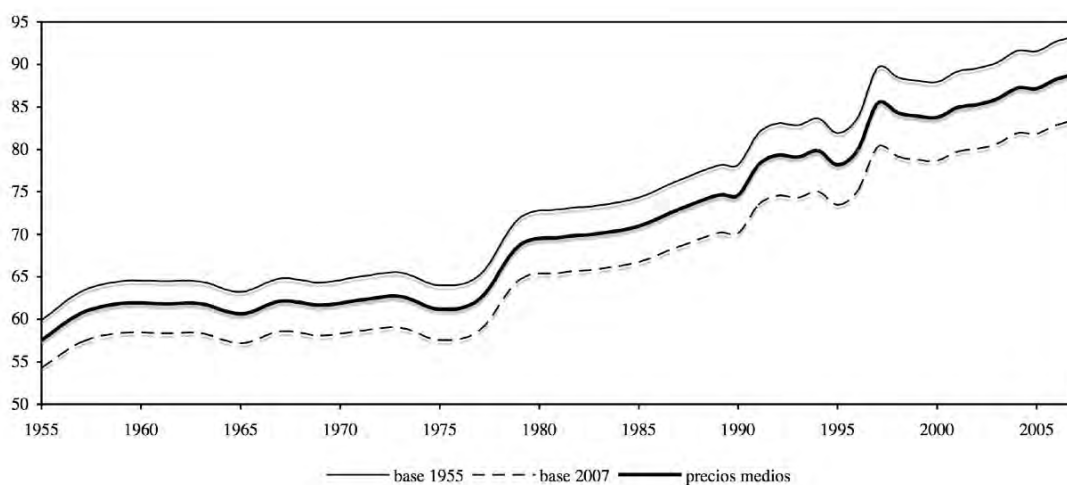


**Gráfico A2.19: Productividad (VAB por empleo a precios constantes)  
deflactado con distintos índices de precios -- continuación**

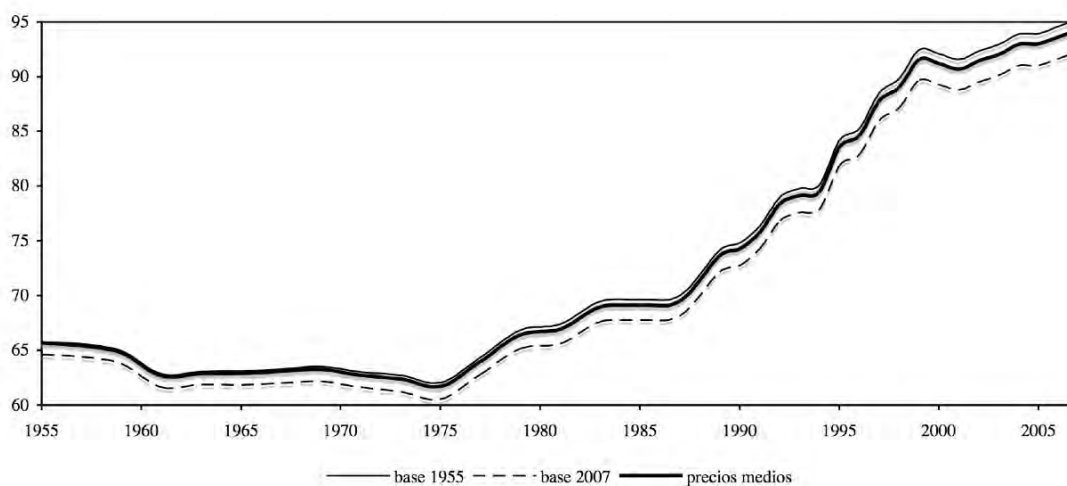
### Valencia



### Extremadura

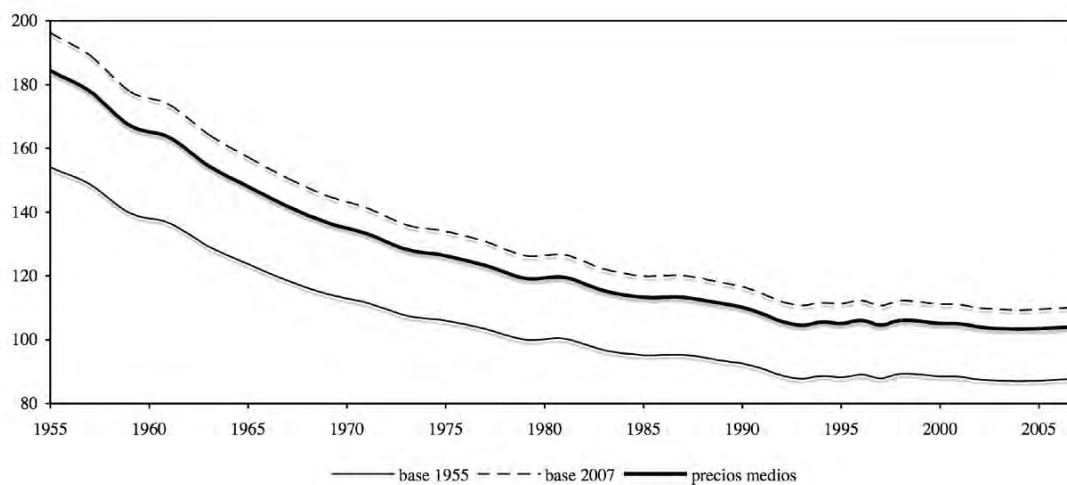


### Galicia

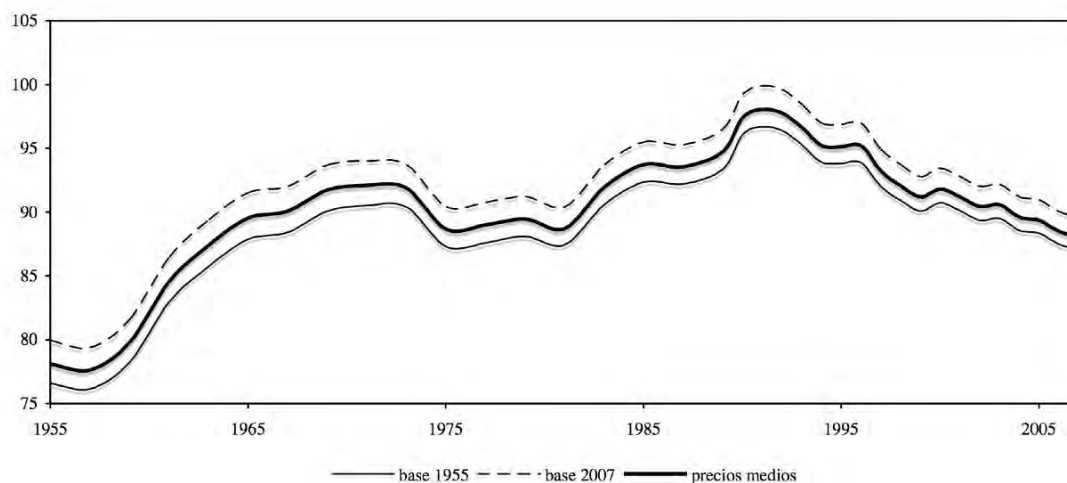


**Gráfico A2.19: Productividad (VAB por empleo a precios constantes)  
deflactado con distintos índices de precios -- continuación**

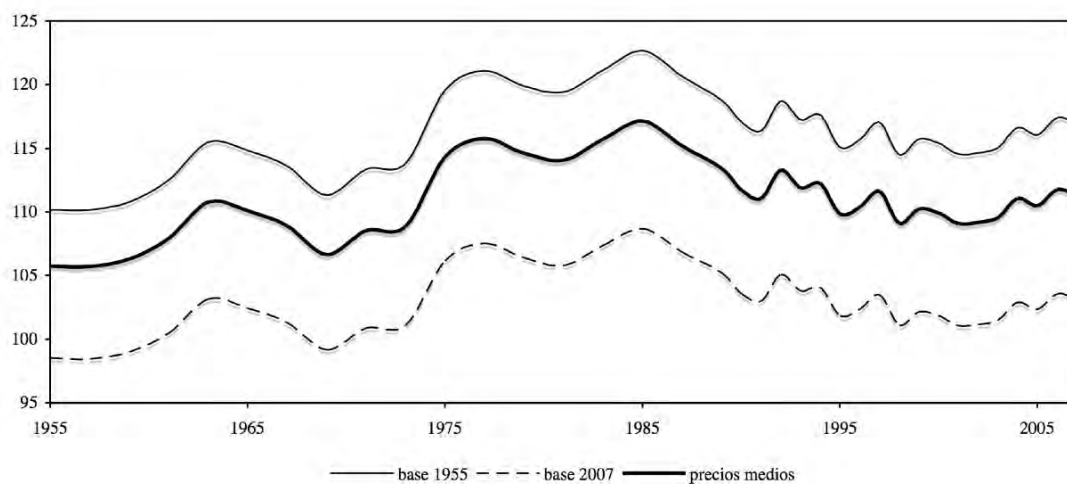
### Madrid



### Murcia



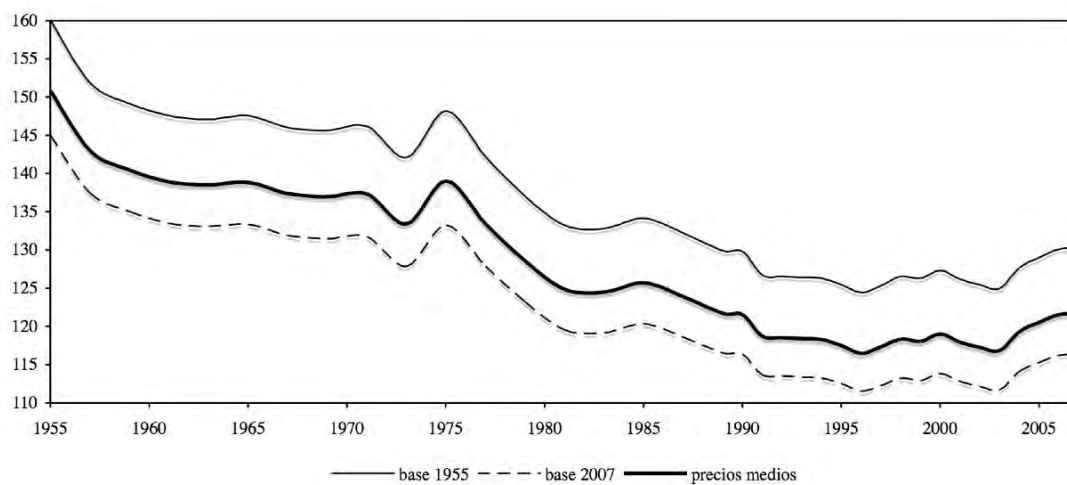
### Navarra



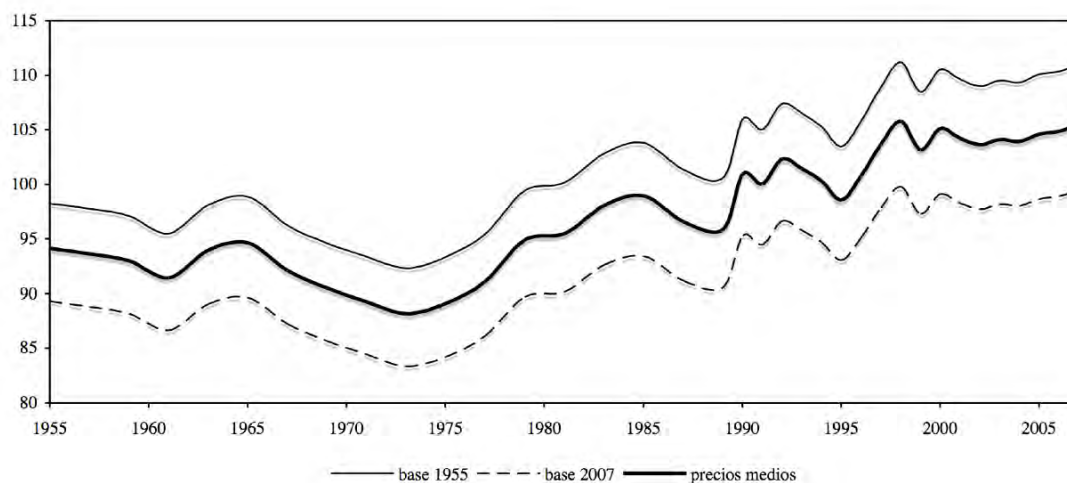


**Gráfico A2.19: Productividad (VAB por empleo a precios constantes)  
deflactado con distintos índices de precios -- continuación**

**País Vasco**



**Rioja**



**Ceuta y Melilla**

